Vítor Manuel Monteiro Borges
Mestre em Ciências Farmacêuticas

Evolutionary dynamics of *Chlamydia trachomatis* genome and identification of molecular patterns of hypothetical protein coding genes

Dissertação para obtenção do Grau de Doutor em Biologia

Orientador: Doutor João Paulo dos Santos Gomes, Investigador Auxiliar, Instituto Nacional de Saúde Doutor Ricardo Jorge

Co-orientadora: Doutora Maria José Gonçalves Gaspar Borrego, Investigadora Auxiliar, Instituto Nacional de Saúde Doutor Ricardo Jorge

Co-orientadora: Doutora Ana Madalena Peres de Drummond Ludovice Mendes Gomes, Professora Auxiliar, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Júri :

Presidente: Prof. Doutora Ana Isabel Nobre Martins Aguiar Oliveira Ricardo

Arguentes: Doutor João André Nogueira Custódio Carriço

Doutora Isabel Maria dos Santos Leitão Couto

Vogais: Doutora Maria Aida da Costa e Silva da Conceição Duarte

Doutora Isabel Maria Godinho de Sá Nogueira

Doutor João Paulo dos Santos Gomes

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Maio, 2015

Vítor Manuel Monteiro Borges
Mestre em Ciências Farmacêuticas

Evolutionary dynamics of *Chlamydia trachomatis* genome and identification of molecular patterns of hypothetical protein coding genes

Agradecimentos

De um modo sincero, agradeço:

Ao meu orientador Dr. João Paulo Gomes, Responsável da Unidade de Investigação do Departamento de Doenças Infecciosas do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), por me ter integrado no seu grupo de investigação, por todo o saber científico que me transmitiu, pelo apoio incondicional, pela liderança, pela permanente disponibilidade e ajuda, pelo encorajamento e amizade.

À minha co-orientadora Dra. Maria José Borrego, Responsável do Laboratório Nacional de Referência das Infecções Sexualmente Transmissíveis do INSA, pelos ensinamentos científicos, pela total disponibilidade, pela constante simpatia, por todo o apreço, apoio e amizade.

A todos os restantes membros da Comissão de Acompanhamento de Tese (CAT), pela discussão científica e por todas as sugestões concedidas. Em especial, à minha co-orientadora Dra. Ana Madalena Ludovice, Professora Auxiliar da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL) por ter aceitado ser minha orientadora, pela simpatia e disponibilidade; e ao Dr. Jaime Mota (FCT/UNL) pela profícua discussão e colaboração científica, apoio e disponibilidade. Agradeço, ainda, ao Professor Miguel Viveiros do Instituto de Higiene e Medicina Tropical (UNL) e à Professora Cristina Lobo Vilela da Faculdade de Medicina Veterinária da Universidade Técnica de Lisboa.

A todos os colegas de investigação do INSA. Em particular, à Alexandra Nunes, ao Carlos Florindo e à Rita Ferreira, pela entreajuda e solidariedade, pela partilha de ideias e pela discussão científica. Foi e tem sido um prazer trabalhar convosco. À Minia Antelo, à Mafalda Sousa-Uva, à Vera Damião e ao Miguel Pinto, pelo companheirismo e colaboração.

A todos os membros do grupo do Dr. Jaime Mota, em especial ao Filipe Almeida pela entreajuda e profissionalismo.

Ao Dr. Miguel Pinheiro da Universidade de St. Andrews (Escócia, Reino Unido), pelos ensinamentos na área da bioinformática, pelo constante encorajamento e suporte, e, consequentemente, pelo seu importante contributo para o desenvolvimento dos trabalhos do último ano de Doutoramento.

A toda a equipa da Unidade de Tecnologia e Inovação do INSA. Em especial, ao seu responsável, Dr. Luís Vieira, pelo seu empenho na implementação e optimização das metodologias de Sequenciação de Nova Geração, as quais foram tão relevantes na fase final deste trabalho.

A todos os demais colegas do INSA, em especial à Dora Cordeiro e à Margarida Dinis, e, sem esquecer, à Dona Albertina Paulino e à Dona Arminda Louro, por todo o apoio e estima.

Ao Instituto Nacional de Saúde Dr. Ricardo Jorge, em particular ao Departamento de Doenças Infecciosas, na pessoa do seu coordenador, Dr. Jorge Machado, por ter disponibilizado todas as condições para a melhor prossecução deste trabalho de Doutoramento.

À Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL), em particular à Professora Isabel Sá Nogueira, Coordenadora do Programa Doutoral de Biologia.

À Fundação para a Ciência e Tecnologia, Ministério da Educação e Ciência (FCT/MEC), pelo financiamento concedido através da Bolsa Individual de Doutoramento (SFRH/BD/68527/2010).

Por fim, agradeço profundamente à minha família e aos meus amigos.

Resumo

A bactéria intracelular obrigatória *Chlamydia trachomatis* é um agente patogénico para o Homem com importante impacto em termos de saúde pública. As estirpes podem ser classificadas em 15 principais serotipos (A a L3) que causam preferencialmente infecções oculares (A-C), infecções genitais (D-K) ou linfogranuloma venéreo (LGV) (L1-L3), contudo a base molecular que justifica o seu distinto tropismo, sucesso ecológico e patogenicidade não está bem definida. A investigação neste organismo exige normalmente a sua cultura em linhas celulares eucariotas, mas é desconhecida a existência de um processo de adaptação laboratorial.

Pretendemos, essencialmente através de estudos de genómica e transcriptómica, investigar os padrões evolutivos subjacentes à adaptação de *C. trachomatis* aos diferentes tecidos humanos, dando ênfase aos perfis moleculares de genes que codificam proteínas com função desconhecida, e procurámos compreender o processo adaptativo inerente à transição de *C. trachomatis* de *in vivo* para *in vitro*.

Os nossos resultados indicam que a bactéria *C. trachomatis* tem uma evolução mediada por eventos de seleção positiva orientada para a adaptação aos diferentes tecidos humanos. Observámos que proteínas que interagem com o hospedeiro têm sido importantes alvos de selecção, nomeadamente proteínas efectoras e da membrana da inclusão, sendo que algumas destas apresentam padrões de expressão relacionados com o nicho biológico que as estirpes infectam. Identificaram-se, ainda, potenciais pseudogenes específicos das estirpes oculares e genes que estarão mais sujeitos a mutações adaptativas associadas ao LGV. Verificámos que a constituição genética de estirpes provenientes do ambiente *in vivo* não é substancialmente afectada pela sua propagação laboratorial a longo termo. Pelo contrário, a introdução de *C. trachomatis* em laboratório pode potencialmente levar à atenuação da sua virulência. De facto, observou-se uma rápida inactivação do gene de virulência CT135, fenómeno que parece ser restrito às estirpes que causam infecções genitais, o que pode ter impacto na interpretação de dados provenientes de estudos que requerem cultura.

Globalmente, os resultados apresentados nesta tese contribuem para a melhor compreensão da evolução adaptativa de *C. trachomatis* e fornecem novos dados sobre o papel biológico de proteínas com função desconhecida. Estes dados poderão servir de base para estudos futuros de funcionalidade focados em clarificar os determinantes das diferenças de tropismo, virulência ou patogenicidade entre estirpes de *C. trachomatis*.

Palavras-chave: *Chlamydia trachomatis*, evolução, mutação, adaptação, tropismo, proteína hipotética, virulência.

Abstract

The obligate intracellular bacterium *Chlamydia trachomatis* is a human pathogen of major public health significance. Strains can be classified into 15 main serovars (A to L3) that preferentially cause ocular infections (A-C), genital infections (D-K) or lymphogranuloma venereum (LGV) (L1-L3), but the molecular basis behind their distinct tropism, ecological success and pathogenicity is not well-defined. Most chlamydial research demands culture in eukaryotic cell lines, but it is not known if stains become laboratory adapted.

By essentially using genomics and transcriptomics, we aimed to investigate the evolutionary patterns underlying the adaptation of *C. trachomatis* to the different human tissues, given emphasis to the identification of molecular patterns of genes encoding hypothetical proteins, and to understand the adaptive process behind the *C. trachomatis in vivo* to *in vitro* transition.

Our results highlight a positive selection-driven evolution of *C. trachomatis* towards niche-specific adaptation, essentially targeting host-interacting proteins, namely effectors and inclusion membrane proteins, where some of them also displayed niche-specific expression patterns. We also identified potential "ocular-specific" pseudogenes, and pointed out the major gene targets of adaptive mutations associated with LGV infections. We further observed that the *in vivo*-derived genetic make-up of *C. trachomatis* is not significantly compromised by its long-term laboratory propagation. In opposition, its introduction *in vitro* has the potential to affect the phenotype, likely yielding virulence attenuation. In fact, we observed a "genital-specific" rampant inactivation of the virulence gene CT135, which may impact the interpretation of data derived from studies requiring culture.

Globally, the findings presented in this Ph.D. thesis contribute for the understanding of *C. trachomatis* adaptive evolution and provides new insights into the biological role of *C. trachomatis* hypothetical proteins. They also launch research questions for future functional studies aiming to clarify the determinants of tissue tropism, virulence or pathogenic dissimilarities among *C. trachomatis* strains.

Keywords: *Chlamydia trachomatis*, evolution, mutation, adaptation, tissue tropism, hypothetical protein, virulence.

Table of contents

Figure index

Table index

List of Abbreviations

ABD	Actin binding domain
ATCC	American Type Culture Collection
BEB	Bayes Empirical Bayes
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary Deoxyribonucleic acid
D	Aspartic acid
d_N	Number of nonsynonymous substitutions *per* nonsynonymous site
DNA	Deoxyribonucleic acid
d_S	Number of synonymous substitutions *per* synonymous site
DUB	Deubiquitinating
CDS	Coding sequence
CEP	Cell Envelope Protein
CPAF	Chlamydial Protease/proteasome-like Activity Factor
EB	Elementary body
FCT/MEC	Fundação para a Ciência e Tecnologia, Ministério da Educação e Ciência
FCT/UNL	Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
FHA	Forkhead associated
G	Glycine
GCIP	Grap2 cyclin D-interacting protein
h	hour
HIV	Human Immunodeficiency Virus
HK	Housekeeping gene
HPV	human papillomavirus
IM	Inner membrane
Inc	Inclusion membrane
Indel	Insertion/deletion
IFN-γ	Interferon-gamma
IFU	Iinclusion-forming unit
IGR	Intergenic region
LGV	Lymphogranuloma venereum
LPS	Lipopolysaccharide
LRT	Likelihood-ratio test
MACPF	Membrane attack complex/perforin
MEGA	Molecular Evolutionary Genetics Analysis
MEM	Minimum Essential Medium

min	minute
MK	McDonald-Kreitman
mRNA	Messenger ribonucleic acid
MSM	Men who have sex with men
MOMP	Major outer membrane protein
MoPn	Mouse Pneumonitis
NCBI	National Center for Biotechnology Information
OM	Outer membrane
ORF	Open reading frame
ORI	Origin of replication
PAML	Phylogenetic Analysis by Maximum Likelihood
PBS	phosphate-buffered saline
PCR	Polymerase chain reaction
pi	post-infection
Pmp	Polymorphic membrane protein
qPCR	Real-time quantitative PCR
RB	Reticulate body
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT	Reverse Transcription
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Secreted protein
SEM	Standard error of the mean
serovar	serological variant
SNARE	Soluble N-ethylmaleimide-sensitive attachment protein receptors
SNP	Single nucleotide polymorphism
sRNA	small RNA
STI	sexually transmitted infection
T3S	Type III secretion
T3SS	Type III secretion system
Tarp	Translocated actin-recruiting phosphoprotein
TCA	Tricarboxylic acid
TepP	Translocated early phosphoprotein
TER	Termination region
TM	Transmembrane
tRNA	Transfer ribonucleic acid
trpR	tryptophan operon repressor gene

trpRBA	Tryptophan synthase operon
Tsp	Tail-specific protease
TSS	Transcriptional start site
WHO	World Health Organization
μ_{bp}	mutation rate *per* base pair *per* replication
μ_{g}	mutation rate *per* genome *per* replication

Notes of the author: thesis organization, format and outline

The main body of this Ph.D. dissertation is based on seven manuscripts (listed below) that are presented as individual chapters (II to VIII). Six of them have already been published (the remaining is submitted for publication at the time this thesis was completed) in peer reviewed international journals, being presented in this thesis essentially as a reproduction of the content that was published. In this context, the chapter presentation order does not perfectly reflect the chronological order of the manuscripts' publication, as several studies were developed simultaneously and the time elapsing between the manuscript submission and publication largely depends on the journal and on the inherent revision requirements. The chapters were organized so that they follow a rational order taking into account the objectives delineated for this Ph.D. work, and in agreement with the association between the scientific subjects addressed in each one, as the results obtained during one study influenced the progress of others, and *vice-versa*. Each manuscript-based chapter is preceded by a title page describing the reference of the publication, the specific contributions of the author of the present Ph.D. thesis, and, where applicable, the add-ons relative to the published content (alterations regarding style adjustments are referred as "minor changes"). Besides these manuscript-based chapters, each one including extensive and specific introduction and discussion sections, the present doctoral dissertation includes a general introduction (chapter I) and a conclusive overview (chapter IX). In brief, each chapter includes the following contents:

Chapter I. This chapter consists of an general introduction that intents to provide the reader with the state of the art in the subjects addressed in the this doctoral dissertation around the major human pathogen *Chlamydia trachomatis*. On behalf of this, it is firstly given a global overview of the major aspects of *C. trachomatis* taxonomy, biology, molecular epidemiology and impact on human health, followed by insights into the evolution, genetic diversity and some already established genotype-phenotype associations. Historical hurdles and ongoing advances on the genomics research in *C. trachomatis* are also reviewed in this chapter. It ends with the description of the main objectives of this Ph.D. project, and includes the specific research questions that drove the investigations carried out on behalf of each chapter.

Chapter II. Borges V., Nunes A., Ferreira R., Borrego M.J., Gomes J.P. 2012. Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation. Journal of Bacteriology, 194:6143-6153.

Chapter III. Almeida F.*, Borges V.*, Ferreira R., Borrego M.J., Gomes J.P., Mota L.J. 2012. Polymorphisms in Inc proteins and differential expression of inc genes among *Chlamydia trachomatis* strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates. Journal of Bacteriology, 194:6574-6585. **These two authors equally contributed to this work.*

Chapter IV. da Cunha M., Milho C., Almeida F., Pais S.V., Borges V., Maurício R., Borrego M.J., Gomes J.P., Mota L.J. 2014. Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system. BMC Microbiology, 14:40.

Chapter V. Borges V., Pinheiro M., Vieira L., Sampaio D.A., Nunes A. Borrego M.J. Gomes J.P. 2014. Complete genome sequence of *Chlamydia trachomatis* ocular serovar C strain TW-3. Genome Announcements, 2:e01204-13.

Chapter VI. Borges V., Gomes J.P. 2015. Deep comparative genomics among *Chlamydia trachomatis* lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation. Infection, Genetics and Evolution, 32:74-88.

Chapter VII. Borges V., Ferreira R., Nunes A., Sousa-Uva M., Abreu M., Borrego M.J., Gomes J.P. 2013. Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics. Infection, Genetics and Evolution, 17:23-32.

Chapter VIII. Borges V., Pinheiro M., Antelo M., Sampaio D.A., Vieira L., Ferreira R., Nunes A., Almeida F., Mota L.J., Borrego M.J., Gomes J.P. *Chlamydia trachomatis in vivo* to *in vitro* transition reveals mechanisms of phase variation and down-regulation of virulence factors *(submitted manuscript)*.

Chapter IX. This chapter provides a global overview of the subjects addressed throughout the chapters, highlighting the main results and conclusions achieved in this Ph.D. dissertation. New research questions raised with this work that can be addressed in the future follow-up of these investigations are also presented.

Considering the dissimilar layouts and in-text reference styles adopted by different journals where the manuscripts were published, all chapters were formatted in a unique style, with all references being cited by sequential numbers (in parentheses) and listed in the "References" section according to the order in which they appear in the text. In this regard, a single section of "References" is presented. Finally, annexes relative to each chapter are also compiled in a last section of this doctoral thesis in a section referred as "Supplemental material". For the sake of simplicity, a link to the online publication is provided for some annexes.

Chapter I

General Introduction

1. General Introduction

1.1. The genus *Chlamydia*: host preference and pathogenicity

The family *Chlamydiaceae* (order *Chlamydiales*, phylum *Chlamydiae*) comprises a single genus, *Chlamydia*, which enrolls 11 currently recognized obligate intracellular bacterial species known to be important pathogens of humans and/or animals: *C. trachomatis*, *C. pneumoniae*, *C. psittaci*, *C. abortus*, *C. felis, C. pecorum*, *C. suis*, *C. avium*, *C. gallinacea*, *C. caviae* and *C. muridarum* (1-3). These species present relevant differences in terms of host range, tissue tropism, and disease pathology and outcomes. The most relevant human pathogens are *C. trachomatis*, which infects exclusively the human being causing mostly ocular and genital infections (4-6), and *C. pneumoniae*, which, besides infecting animals such as horses, marsupials or frogs, has been implicated in acute human respiratory infections, and also in chronic diseases such as obstructive pulmonary disease, atherosclerosis or even type 2 diabetes (7-10). Other species which may have impact on human health due to their potential for zoonotic transmission are *C. psittaci* and *C. abortus*, and to a lesser extent *C. felis* (11). *C. psittaci* is the pathogenic agent of an avian respiratory disease (avian chlamydiosis or psittacosis), but severe cases of human respiratory infections have been described (12-15). On the other hand, *C. abortus* infects a wide range of animals (such as sheep, cattle or goats) being mostly known as a major cause of ovine enzootic abortion. Nevertheless, pregnant women are also susceptible to *C. abortus* infections, as this pathogen targets the placenta, potentially leading to abortion (16-18). *C. felis* is a causative agent of infections in the upper respiratory tract and eyes of cats and also displays zoonotic potential as there have been some reports of human conjunctivitis after contact with infected cats (19-21). Although the remaining species of the genus *Chlamydia* are not of recognized direct concern to human health, they may cause a wide diversity of animal diseases with significant economic impact. This particularly stands for *C. pecorum*, which is a pathogen of a wide range of animals (e.g., cattle and other ruminants, sheep, swine and koala) (22-24), and for *C. suis*, which is a natural pathogen of pigs (25,26). Of note, the recent description of two new species, *C. avium* and *C. gallinacea*, displaying ability to infect birds (1) and potentially cause respiratory disease (27) implies that their zoonotic potential and economic impact cannot be discarded, and that the etiology of cases of avian chlamydiosis needs to be revisited (27). Finally, *C. caviae* may cause conjunctivitis and genital tract infections in guinea pig (28), whereas *C. muridarum* [the most genetically related species to *C. trachomatis* (29)] is the natural mouse chlamydial pathogen, being able to cause pneumonitis, but also infections in the urogenital tract in mice (30). These two species have been used to model chlamydial urogenital infections. In particular, the model "mouse - *C. muridarum*", which is also used as a surrogate for studying respiratory chlamydial infections, is the most popular and extensively used model for the acquisition of knowledge about the biology and immunopathology of the important human pathogen *C. trachomatis* (31-33).

1.2. *Chlamydia trachomatis*

1.2.1. Molecular epidemiology, tissue tropism and impact on human health

C. trachomatis strains have been traditionally classified into 15 main serological variants (serovars) based on the differential immunoreactivity of its major outer membrane protein (MOMP) or based on the polymorphism of the gene *ompA* (which codes for MOMP) (34). The serological profiles match the corresponding *ompA* genotypes, so that the "serovar" designations still persisted in the era of the molecular typing. Although both the serological relatedness of MOMP serovars and *ompA* phylogeny fail to reflect either the tropism of strains, or its pathogenicity and clinical prevalence (34-40), the applied nomenclature strongly links to those properties. In fact, strains from serovars A to C typically infect the epithelial cells from the conjunctival mucosa (and are normally named as "ocular strains") potentially leading to trachoma, a chronic eye disease that may lead to irreversible blindness (6,41-43). In turn, strains from serovars D to K are usually associated with localized infections of the epithelial surface of the urogenital (and also ano-rectal) mucosa [being commonly designated as "(uro)genital" or "epithelial-genital" strains], and are the major cause of bacterial sexually transmitted infections (STIs) worldwide (4,44). Finally, strains from serovars L1-L3 are the causative agents of lymphogranuloma venereum (LGV), which is an inguinal syndrome normally characterized by genital ulceration and painful inguinal lymphadenopathy (inguinal buboes) (45,46). The capacity of these so-called "LGV strains" to cause that particular clinical presentation (the typical bubonic LGV) relies on their ability to infect mononuclear phagocytes (upon genital or rectal entry) and disseminate to regional lymph nodes (46,47). Since 2003, an atypical LGV clinical presentation characterized by severe ulcerative proctitis (anorectal syndrome) and primarily caused by L2b strains has emerged in Europe and North America (5,48,49). Of note, as both A-C and D-K strains are normally epitheliotropic and mucosae-restricted, they have been historically grouped in a specific biovar designated as "trachoma biovar", in opposite to the "LGV biovar" that is composed by the LGV strains (50). Additionally, despite being an atypical outcome, strains from genital serovars can occasionally be detected in the ocular tract and *vice-versa* (51,52).

Considering its ample tissue tropism and huge potential to cause a wide range of human diseases, the pathogen *C. trachomatis* has been considered as a global health problem, as revealed by its clinical relevance, prevalence and geographical distribution in the worldwide picture of infectious diseases. In fact, trachoma is the world’s leading cause of preventable infectious blindness (6,41,42), mostly affecting poor communities with little or no healthcare from Latin America, Sub-Saharan Africa or Asia (where serovar A and B strains seem to prevail), and some remote Australian communities (where serovar C is common) (11,52,53). In 2012, the World Health Organization (WHO) estimated that more than 40 million people in over 50 countries are affected by trachoma (over 8 million at risk of blindness) (54); however, the prevalence is declining as a result of the 1998 initiative implemented by the WHO - The Global Elimination of Trachoma by 2020 (55). In turn, STIs by *C. trachomatis* genital strains are commonly asymptomatic (in about 70% of women and 50% of men) or cause cervicitis or urethritis (56-58). Untreated infections may progress to epididymitis (men), or lead to pelvic

inflammatory disease and other severe complications, including chronic pain, ectopic pregnancy and infertility (women) (57-59). Over 100 million cases of STIs due to *C. trachomatis* are believed to arise annually (44), and, despite being effectively treatable with antibiotics (azithromycin and doxycycline are first choices), cases of antibiotic resistance could be increasing (58). Most worldwide cases of epithelial-genital tract infections in heterosexual populations have been found to be related to serovars E and F, whereas serovar D, G, and J seem to be more frequently detected in non-LGV rectal infections in men who have sex with men (MSM) (60-65). Noteworthy, the ongoing epidemics of LGV is also raising special concern to health authorities in the Western World. In fact, the underlying serovar L2b ano-rectal infections, which mostly affects MSM [usually co-infected with the human immunodeficiency virus (HIV) and other sexually transmitted pathogens] (66,67), may progress to the typical bubonic LGV (5,67-69), and cases of infections in woman (66,70-72) or treatment failure (doxycycline is the antibiotic of choice) have been described (5,73,74). Finally, although the hallmark of *C. trachomatis* is its ability to cause disease affecting three distinct human tissues (ocular mucosa, genital mucosa and lymph nodes), other diseases have been associated with *C. trachomatis* infections, particularly the *Chlamydia*-related reactive arthritis (75-78). Increasing evidence has been also implicated *C. trachomatis* infection as a risk factor for HIV acquisition (58,79-81) or human papillomavirus (HPV) infection (82,83), and a pro-carcinogenic role of *C. trachomatis* cannot be discarded (84-86).

1.2.2. Biology: a specialized life-cycle of host-cell dependence and manipulation

Similarly to all members of the family *Chlamydiaceae*, *C. trachomatis* has a Gram-negative type cell wall, although peptidoglycan was only very recently identified (87), and follows a unique biphasic developmental cycle characterized by the interconversion between two morphologically distinct forms: an infectious form, the elementary body (EB); and a noninfectious replicating form, the reticulate body (RB) (88-91). EBs are small (~0.3 µm diameter), more metabolically constrained than RBs, and contain a central and dense nucleoid and a rigid cell wall that enable them to transitorily resist and survive in the extracellular environment (91-94). On the other hand, RBs, which are larger than EBs (~1 µm diameter) and osmotically fragile (due to the lack of the highly cross-linked outer membrane typical of EBs), contain relax chromatin, and are highly metabolically active as they sustain critical energetically costly functions, such as uptake and transport of nutrients, or abundant protein synthesis (91-94). The life-cycle (Figure 1.1) of about 30 to 72 hours comprehends the following major steps: *i)* EB attachment and entry into the target eukaryotic cells, and rapidly establishment of the chlamydial parasitophorous membrane-bound vacuole, termed inclusion; *ii)* EBs differentiation into RBs inside the inclusion, which continuously expands while RBs are replicating by binary fission (doubling times of about 2 to 4 h); *iii*) RB to EB differentiation in an asynchronous manner (i.e., whereas some RBs continue to replicate, others differentiate back into EBs); and, finally, *iv*) bacterial release by host cell lysis or extrusion, and dissemination of infectious EBs to target neighboring cells (88-91,95). This normal developmental cycle can be disturbed under particular adverse growth conditions. *In vitro*, it has been shown that nutrient starvation, the presence of penicillin, iron or

interferon-gamma (IFN-γ) lead to a state of persistence, where the bacteria remains viable but not cultivable, being characterized by morphologically enlarged, aberrant and non-replicative RBs (96,97).

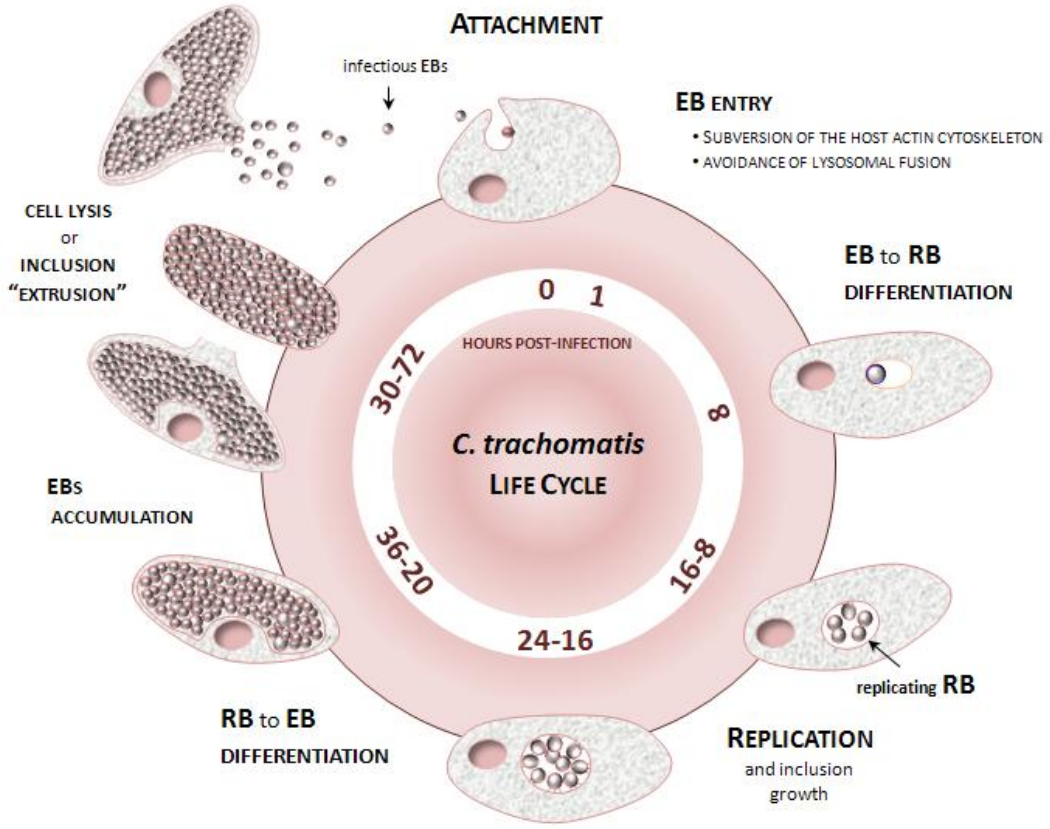


Figure 1.1. Schematic representation of the *C. trachomatis* developmental cycle.

C. trachomatis undergoes a temporal program of gene expression essentially marked by early, mid-cycle and late genes, which perfectly links to the progression of the developmental cycle (98-101). Early genes, including some immediately early (< 1 h post-infection), are transcribed within about three hours of EB uptake, and are believed to be important for establishing the intracellular infection. Mid-cycle genes, which constitute the large majority of the chlamydial genes, code for proteins essentially involved in bacterial replication and in subversion of host-cell functions for intracellular growth maintenance. Finally, late genes are first expressed or up-regulated at the end of the developmental cycle, with some of them (called very-late or "tardy" genes) encoding proteins that have been shown to be pre-packaged into EBs for playing relevant roles in the subsequent round of host-cell invasion and infectivity (98-103). Genes from the three major temporal classes can be transcribed by the major chlamydial RNA polymerase sigma factor σ^{66} [the homolog of *Escherichia coli* main σ factor (σ^{70})], although *C. trachomatis* also encodes two alternative sigma factors, σ^{28} and σ^{54} (104-106). Key bacterial factors playing central roles in the chlamydial development and/or gene regulation include: *i*) proteins implicated in the host-cell adherence, such as the cysteine-rich outer membrane

protein OmcB (107,108) and polymorphic membrane proteins (Pmps) (109-112); *ii*) the histone-like proteins HctA and HctB that mediate the chromosomal condensation in the differentiation of RBs and EBs (113,114); *iii*) porins that ensure nutrient transport (e.g., MOMP and PorB) (115,116); or *iv*) global transcriptional regulators, such as the small RNA (sRNA) IhtA, which was suggested to control the timing of RB to EB transition (117,118), and the repressor of late genes EUO (119,120).

Throughout all life-cycle, *C. trachomatis* intensively interacts and exploits the host cell by subverting critical cellular functions (such as, host cytoskeleton assembly, cytokinesis, apoptosis, nutrient transport, membrane trafficking pathways or immune responses), thus ensuring its specialized intracellular growth (121-126). This remarkable capacity to manipulate the host is essentially carried out by host-interacting proteins that are translocated into the host cytosol (so called “effectors”) or are localized into the inclusion membrane (103,127,128). The latter, designated as inclusion membrane (Inc) proteins, are inserted into the vacuolar membrane likely due to the possession of an amino acid bilobed hydrophobic motif (129-132). With the exception of some relevant translocated proteases, such as the chlamydial protease-like activity factor (CPAF) and the tail-specific protease (Tsp) (likely secreted via a sec-dependent pathway) (133-136), most of the so far described *C. trachomatis* effectors and Incs were found to be transported by using a type III secretion (T3S) system (103,127,129-132,137,138). This mechanism is widely used by many bacterial pathogens to manipulate eukaryotic host cells by injecting virulence proteins into their cytosol and membranes (139). Two *C. trachomatis* T3S substrates that have been raising special attention in the chlamydial research field include: the translocated actin-recruiting phosphoprotein (Tarp), an early-secreted effector known to play a critical role during the host-cell invasion (140,141); and IncA, an Inc protein involved in the subversion of the intracellular trafficking by inducing the homotypic fusion of intracellular inclusions and protecting them from the endolysosomal pathway (142-145). Although the biology of *C. trachomatis* clearly supports its capacity to promote virulence, the factors underlying the differential tissue tropism and disease severity displayed by distinct *C. trachomatis* strains are not completely understood. Nevertheless, during the complex host-pathogen "arms race" that takes place during *in vivo* infections, it is expected that variable host-related factors (e.g., genetics and/or immunological) (146,147) and dissimilar genetic backbones of the *C. trachomatis* infecting strains determine the type and fate of each infection.

1.2.3. Evolution, genetic diversity and genotype-phenotype associations

Due to their obligate intracellular lifestyle, all members of *Chlamydiaceae* family have reduced genomes (1-1.2 Mb) (2,11). In fact, these organisms suffered reductive evolutionary processes characterized by the loss of genes that became non-essential in the intracellular habitat, mainly those involved in biosynthetic functions required to produce metabolites that are also available in the host cell cytoplasm (115,148,149). This genome contraction may have also relied on the consequent reduction of the effective population sizes, implying evolutionary bottlenecks (and reduced rates of gene flux) towards a less efficient capacity to purge slightly deleterious mutations from the populations (148), in a phenomenon commonly known as Muller's ratchet (150-153). Nevertheless, the full-

sequencing of *C. trachomatis* genome in 1998 (115) revealed that this bacterium likely reached the final stages of its reductive evolution (148). In fact, apart from its ~7,5 kb plasmid (identified in 1980) (154), *C. trachomatis* has a small circular chromosome (~1 Mb) (%G/C of ~41,3) marked by a high coding density (~90%) and expectedly few pseudogenes and non-essential genes (115). It also reveals no evidence of the occurrence of recent genetic exchange with other non-chlamydial bacterial species, neither transposons or prophages, nor any evidence for the recent acquisition of novel genes (155). Globally, its genome structure enrolling about 900 genes is likely reaching stability, contrarily to other organisms sharing a similar reductive evolution, namely *Rickettsia prowazekii*, whose chromosome contains a low coding content (about 75%) and multiple remnants of ancient genes that are potentially under the elimination process (149). The minimal genome of *C. trachomatis*, together with its extraordinary capacity to promote complex interactions with the host cellular machinery, further implies that it encodes a high density of virulence-related proteins (115,156). On this behalf, it is remarkable that ~32% of the *C. trachomatis* genome enrolls genes encoding proteins with unknown function (generally called "hypothetical proteins"), where only ~4% of them are believed to occur in other bacteria ("conserved hypothetical proteins") (115). Although the proportion of hypothetical proteins is overestimated because experimental proofs are not being systematically updated in genome annotations, cumulative data have been pointing relevant biological roles for these uncharacterized proteins in the complex cascade of host-pathogen interactions that take place during *C. trachomatis* infections. In fact, they seem to be overrepresented by host-interacting proteins, namely effectors and Incs (103,132), and some of them were found, for instance, to encode strong antigens (157-162) or to be highly polymorphic among *C. trachomatis* strains (46,163-165). Therefore, studies aiming to assign and characterize their molecular profiles may certainly contribute to better understand the biology and pathogenesis of *C. trachomatis*, particularly the underlying features that justify the dissimilar tropism, virulence and ecological success displayed by different strains.

Studies focusing on the genetic diversity among *C. trachomatis* strains from the three “disease groups” (ocular, genital and LGV) revealed that they have a similar core and pan-genome, with an extremely conserved gene order and a remarkable nucleotide sequence similarity > 98% (37,46,115,165,166). While these findings discard horizontal gene transfer as a major evolutionary operator for generating diversity within the *C. trachomatis* species, they immediately pointed that the phenotypic dissimilarities observed between and within strains from each disease group should rely on few genes present (or activated) in some strains but not in others, and/or on few nucleotide differences or small insertions/deletions (146,155,167). Concordantly, subtle genomic variations among *C. trachomatis* ocular strains were suggested to mediate virulence dissimilarities upon infection of non-human primates (164). It has also been demonstrated that *C. trachomatis* may undergo intra- and intergenic homologous recombination among strains exhibiting the same or dissimilar tissue tropism, where the regions flanking the gene *ompA* are thought as hotspots for recombination (37,38,168-175). However, its obligatory intracellular lifestyle constitutes a barrier to cell-to-cell contact and thus, recombination is thought to be privileged between strains with tropism for the same tissue upon mixed infections, which are believed to occur at a frequency of 1% (146,155,167). According to computational inferences of the frequency and relative weight of recombination and mutation events

(173,174,176), point mutations are likely the primary evolutionary driving force for strains' diversification, whereas recombination events have also certainly played a relevant role in the never-ending adaptive process of bacterial fitness improvement (11,167).

Collectively, this scenario of genetic diversity within the *C. trachomatis* species sustains that the detection of genes with molecular signatures concordant with the different tissue tropism, ecological success or pathogenic differences among same-niche infecting strains is of crucial relevance. In fact, the search for the biological basis behind such phenotypes has been a permanent goal of the research on *C. trachomatis*. As a result of this, there are already some well-established genetic features that distinguish *C. trachomatis* strains with distinct tissue tropism (146,167,177,178), particularly the differences enrolling the tryptophan synthase operon (*trpRBA*) and the cytotoxin locus (51,179,180). In this regard, whereas the typical ocular strains display a non-functional *trpRBA* operon, as a result of the disruption of one of the genes encoding the alpha and beta subunits of the tryptophan synthase (CT171/*trpA* and CT170/*trpB*, respectively), the genital strains harbor a predicted functional enzyme, which enable them to biosynthesize tryptophan from exogenous sources, like the indole (51,180). Given that indole is abundantly produced by competing flora on genital tract, such ability is thought to function as an important mechanism for the survival of genital strains under tryptophan-limiting conditions caused by host IFN-γ immune response (a primary anti-chlamydial human immune response), which acts by induction of the tryptophan-catabolizing enzyme indoleamine-2,3-dioxygenase (51,180,181). Regarding the gene encoding the *C. trachomatis* cytotoxin (CT166), which is involved in promoting the disassembly of cytoskeleton actin filaments during bacterial internalization (''cytophatic effect'') (51,182-184), it may be active, truncated or deleted among *C. trachomatis* strains. It is known that the epithelial-ocular and epithelial-genital strains retain remnants of a larger ancestral cytotoxin gene, which is still intact in other *Chlamydia* species (11,177,185,186). However, whereas oculotropic isolates only retain an UDP-glucose binding domain, and thus are non-cytotoxic, the typical epithelial-genital isolates harbor a functionally active CT166, since they encode an additional functional glycosiltransferase domain (179,187). In fact, as seen for the well-known microbial toxin homologs, like the clostridial glucosyltransferases, both domains are expectedly required for the enzymatic activity of CT166 in *C. trachomatis* (179,185,187,188). In the case of the LGV strains, CT166 is completely deleted (179). Finally, it is worth noting that *C. trachomatis* genomes contain a variable in size ~20-25 kb region (locus ~CT152-CT176) called "plasticity zone", which is known to display not only a high level of genetic variation between *C. trachomatis* and other chlamydial species, but also important genetic differences within the *C. trachomatis* species, including those occurring in the *trpRBA* and the cytotoxin loci (167,189). Thus, although significant genotype-phenotype associations have been established, they do not completely explain the molecular basis underlying the ability of different *C. trachomatis* strains to infect different human tissues, neither why some strains are preferentially mucosotropic while others are lymphotropic, nor the dissimilar prevalence and pathogenicity of strains infecting similar niches. While advances in this field have unsurprisingly been slowed down due to research challenges associated with the obligate intracellular nature of *C. trachomatis*, a promising future for chlamydial functional genomics research is coming.

1.2.4. Genomic research: historical hurdles and ongoing advances

Animal models have been central for the study of *C. trachomatis* infections, and thus, for the advances towards the vaccine development (33). Still, none of the available animal models, including the *C. trachomatis* mouse model (used to study genital infection) or the *C. trachomatis* non human primates model (used to study ocular infection), are believed to perfectly reproduce the immune responses and pathogenesis that occurred during human *C. trachomatis* infections (33,190,191). Besides the studies focusing the *C. trachomatis* immunopathology, crucial advances on *C. trachomatis* research have taken advantage of the ability of *C. trachomatis* to be cultured using eukaryotic cell lines. Nonetheless, more than a half century since T'ang and et al (192) successfully isolated *C. trachomatis* through the inoculation of the yolk sack of an embryonated hen egg, the obligate intracellular life-style and the complex biphasic developmental cycle of this bacterium still make its laboratory maintenance as a laborious procedure. In fact, despite the recent advances that have been recently done towards the development of host-cell free (axenic) growth systems (94), *C. trachomatis* only proliferates inside eukaryotic cells (the HeLa229 and McCoy are the most common cell lines used). Moreover, *C. trachomatis* purification procedures, which will ultimately concentrate it in a viable form, generally involve several time-consuming centrifugation steps (193). Also, contrarily to studies conducted with free-living bacteria, there is always the risk of contaminating experiments with host cell proteins, DNA or RNA, although important progresses were recently done towards the whole-genome sequencing of *C. trachomatis* directly from clinical samples (194-197) or the enrichment of the ratio of pathogen RNA to host RNA (101,102). Nonetheless, as *C. trachomatis* culture is mandatory for multiple research purposes, hundreds of strains have been isolated from clinical specimens. Still, most studies (either *in vitro* or *in vitro*) have been carried out using "prototype" (or "reference") strains, which were isolated up to 60 years ago and that have been laboratory maintained since then (63). In this regard, the use of these laboratory-propagated strains has been frequently questioned in *Chlamydia* research, as a result of the assumption that culture propagation may affect the genomic make-up of strains, or even, its transcriptomic or virulence traits (32,195,198-201). Therefore, research is required to characterize the adaptive process underlying the *in vivo* to *in vitro* transition of *C. trachomatis*, and also the impact of the long-term *in vitro* propagation of this bacterium. Despite these culture-associated issues, the advent and continuous improvement of techniques for tissue culture growth or bacterial isolation have made tremendous advancements in *C. trachomatis* research. For instance, isolation of null mutants of the gene CT135 through plaque formation assays followed by *in vivo* experiments allowed to further implicate this uncharacterized gene as a critical virulence factor (202,203).

Another major hindrance for the study of *C. trachomatis* has been the historical lack of tools for its genetic manipulation and transformation, which have hampered the molecular dissection of the virulence factors involved in pathogenesis (204). However, seventeen years after a first promising attempt to genetically transform *C. trachomatis* in 1994 (205), Clarke and co-workers (206) described a workable plasmid-based transformation procedure using antibiotic selection and calcium chloride treatment to render competence. Two years later, in 2013, this procedure allowed the transformation

of plasmidless strains with plasmid shuttle vectors with deletions in each of the eight *C. trachomatis* plasmid coding regions (207). This study revealed the plasmid-encoded Pgp4 as a transcriptional regulator of multiple virulence genes, and consolidated the role of the plasmid in the *C. trachomatis* pathogenesis. In fact, plasmid deficient strains, which occurrence both *in vivo* or *in vitro* is considered to be a rare phenomenon (208-211), are highly attenuated *in vivo* (203). Meanwhile, multiple forward and genetic approaches were developed and the list of plasmid-based gene delivery vehicles is progressively growing (212-220), demonstrating the genetic tractability of *Chlamydia*. The most remarkable approach combines chemical mutagenesis, whole genome sequencing and a system of DNA exchange within infected cells (217,220,221).These advances are considered major steps towards the elucidation of the biology/pathobiology of *C. trachomatis*. Nevertheless, so far, these techniques were mostly used to confirm previously established genotype-phenotype associations (as proof of principle), and their routinely application is only available in very few laboratories.

Bioinformatic analyses have also been a powerful tool to predict virulence factors of intracellular pathogenic bacteria, taking advantage of homology features with proteins in other bacteria, namely model organisms such as *Escherichia coli* and *Bacillus subtilis* (222). Even so, as *Chlamydiaceae* are phylogenetically deeply separated from other eubacteria, the inferences of gene function from the sequence are not straightforward because the level of homology with known proteins may be considerably low (223). On the other hand, the use of comparative genomics to identify inter and intra-species genetic differences has the potential to generate important research questions for further studies aiming to link genotype with phenotype (e.g., functional or epidemiological studies) (177). In this regard, the release of dozens of *C. trachomatis* genomes in 2012 (37) opened avenues for further analyses aiming to comprehend the mutational dynamics behind the strains' divergence, to establish novel and/or strongest genotype-phenotype associations and, ultimately, to identify key genes involved in pathogenesis.

1.3. Aims and General Research Plan

According to the described background, the major aims of this thesis can be pointed out as follows:

- To elucidate the evolutionary patterns underlying the separation of *C. trachomatis* as a species, and its adaptation to the different human tissues (eye, genitalia and lymph nodes). We gave special emphasis to the identification of proteins whose genomic and transcriptomic profiles sustain their contribution to the differential tissue tropism and/or the dissimilar disease outcomes and clinical prevalence of *C. trachomatis* strains.

- To understand the adaptive process behind the *C. trachomatis in vivo* to *in vitro* transition by assessing the impact of introducing and long-term propagating this bacterium *in vitro* on its genome and transcriptome.

In order to pursue these general objectives, we carried out several studies (each one constituting one distinct chapter of this thesis) attempting to answer to the following main research questions:

i) Which are the mutational trends behind *C. trachomatis* speciation and strains´s radiation? Which was the role of selection on these evolutionary processes? Which type of genes were particularly targeted by selection? (**Chapter II**);

ii) Does the polymorphism and/or the differential expression of putative virulence genes (mostly substrates of the virulence-associated T3S system) correlate with the tissue tropism of strains? (**Chapter III and IV**);

iii) To what extent the complete genome sequencing of a *C. trachomatis* serovar C strain may increase our knowledge of the genetic variability among ocular strains and, thus, our global understand of the pathogen factors responsible for the diverse trachoma-related clinical outcomes? (**Chapter V**);

iv) Which are the global trends of genetic variability among *C. trachomatis* strains capable to infect macrophages? Are there L2b-specific genetic features that may enable the epidemic L2b strains to exhibit wider tropism and transmission skills than strains from the remainder LGV serovars? (**Chapter VI**);

v) During the *C. trachomatis in vivo* to *in vitro* transition, which are the genomic and transcriptomic events underlying the adaptation? Is the adaptive dynamics dependent on the tissue-tropism of the infecting strains? (**Chapter VII and VIII**).

In this context, this doctoral dissertation includes cumulative reports from several investigations (which can be read separately) continuously marked by insights into the evolutionary dynamics of *C. trachomatis* genome and the molecular patterns of genes encoding hypothetical proteins, as reflected in the title of the present Ph.D. thesis.

Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation

Manuscript with minor changes published in

2012, Journal of Bacteriology, 194:6143-6153.
Borges V., Nunes A., Ferreira R., Borrego M.J., Gomes J.P.
Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation.
(http://www.ncbi.nlm.nih.gov/pubmed/22961851)

Personal contribution

VB contributed to the design of the study, performed most of the experimental work and bioinformatics analyses, interpreted data and wrote the manuscript.

2. Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation

2.1. Abstract

On behalf of host–pathogen arms race, a cutting-edge approach for elucidating genotype-phenotype relationships relies on the identification of positively selected *loci* involved in the pathoadaptation. We studied the obligate intracellular bacterium *Chlamydia trachomatis*, for which same-species strains display a nearly identical core and pan genome, while presenting a wide range of tissue tropism and ecological success. We aimed to evaluate the evolutionary patterns underlying species separation (divergence) and *C. trachomatis* serovar radiation (polymorphism), and to establish genotype/phenotype associations. By analyzing 60 *Chlamydia* strains, we detected traces of Muller's ratchet as a result of speciation, and identified positively selected genes and codons hypothetically involved in infection of different human cell types: columnar epithelial cells of ocular or genital mucosae, and mononuclear phagocytes; and also, events likely driving pathogenic and ecological success dissimilarities. In general, these genes code for proteins involved in immune response elicitation, proteolysis, subversion of host-cell functions, and also proteins with unknown function. Several genes are potentially involved in more than one adaptive process, suggesting multiple functions or a distinct *modus operandi* for a specific function, and thus should be considered as crucial research targets. Additionally, six out of the nine genes encoding the putative antigens/adhesins polymorphic membrane proteins seem to be under positive selection along specific serovars, which sustains an essential biological role of this extra-large paralogues family in chlamydial pathobiology. This study provides insight into how evolutionary inferences illuminate ecological processes such as adaptation to different niches, pathogenicity, or ecological success driven by arms races.

2.2. Introduction

Genomic changes of microbial pathogens are directly linked to the evolutionary arms race that takes place between microbe and host during the infectious process, as a result of the antagonistic interaction, and they are a consequence of polymorphisms accumulated after selective pressure from the host's inflammatory or immune response (224). However, the majority of coding genes present a higher number of synonymous rather than non-synonymous substitutions, which indicates that purifying selection is operating to preserve the current function and structure of the protein, and only a small fraction of the genes are expected to be positively selected where diversification is favored through increased fitness (225). In order to understand the evolutionary forces that act on gene variation, major challenges are to identify *loci* that might have been under selection, and to determine the type of natural selection that has influenced their evolutionary history (226). In the field of infectious diseases, site-specific inferences regarding positive selection on *loci* involved in drug resistance (227) or in the interaction with the host immune system have been proposed as complementary approaches for the development of vaccines against HIV and other viruses (228), and also to predict the evolution of virulent strains of the influenza virus (229). Also, it has been shown that

core genes are equally subjected to positive selection as pathogen specific accessory genes (230), suggesting that blind genomic-scale analysis should be performed.

For a species such as *Chlamydia trachomatis* with a wide range of tissue tropism and ecological success, but presenting a nearly identical core and pan genome, and a DNA sequence similarity of > 98% (37), the few existing polymorphisms are expected to be extremely informative of the adaptive evolution process. However, an excess of nonsynonymous substitutions alone is not sufficient to invoke positive selection, as it requires an increase in fitness caused by the corresponding amino-acid replacement. Otherwise, it may represent the accumulation of slightly deleterious mutations (not severe deleterious as these will not become fixed because they render their bearers non-viable) to the pathogen on behalf of the Muller's ratchet theory (150,151). This is predicted to operate in intracellular replicating bacteria (as *C. trachomatis,* which replicates within a host vacuole named inclusion) that are subject to recurrent bottlenecks and replicate in small populations, with little opportunity for recombination and few back or compensatory mutations (153). Although it was recently shown (37) that recombination events affect much more chromosome regions than previous suspected in *C. trachomatis*, the frequency and the relative weight of recombination and mutation calculated for this pathogen ($\rho/\theta < 0.07$ and $r/m < 0.71$, respectively) (174,176) indicates the point mutation events as the major evolutionary driving force.

In the present study, we used comparative genomics over 59 *C. trachomatis* strains (comprising all serovars) to clarify the mutational dynamics underlying both the separation of *C. trachomatis* as species, and the pathoadaptation driven by arms race. We identified positively selected genes and codons that are hypothetically involved in the evolutionary adaptation of *C. trachomatis* serovars to different cell types: mucosal cells from the eye conjunctiva (responsible for trachoma) (serovars A-C), from the genitalia (primarily yielding cervicitis) (serovars D-K), and mononuclear phagocytes (yielding invasive diseases such as hemorrhagic proctitis and suppurative lymphadenitis) (serovars L1-L3). Finally, we also detected positive selection events likely driving pathogenic and ecological success dissimilarities.

2.3. Materials and Methods

2.3.1. *C. trachomatis* strains, cell culture and DNA extraction

The present study encompasses data from 59 *C. trachomatis* strains and the *Chlamydia muridarum* Nigg strain (also called, Mouse Pneumonitis strain – MoPn) (listed in the Supplemental Table S2.1). These include *in silico* data from recently analyzed fully-sequenced *C. trachomatis* strains (37) and eight historical prototype strains (Ba/Apache-2, C/TW3, F/IC-Cal3, G/UW57, H/UW43, I/UW12, J/UW36 and K/UW31) in order to enroll all the 15 major serovars. Those additional eight strains were propagated in HeLa 229 cell monolayers and at 48 to 72 hours post-infection, cells were harvested and a bacteria-enriched pellet was obtained and resuspended in 200 µl of PBS, as previously described (231). DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen) according to manufacturer's instructions, and stored at -80 ºC until use.

2.3.2. Selection of *loci* and sequencing

Based on available *in silico* full-genome sequence data, we searched for polymorphic genes among *C. trachomatis* strains through the progressiveMauve algorithm (232) of the Mauve software v2.3.1. A detailed evaluation of polymorphism of each *locus* was further performed by using Lasergene® 9.0. (DNASTAR, Madison, Wisconsin, USA) and MEGA5 (233). Chromosome *loci* revealing an extremely low polymorphism were discarded from the present analysis as their use would hamper the accurate application of likelihood tests. We ended up with 75 top-ranked polymorphic genes (listed in the Supplemental Table S2.2). These were categorized according to their functional role, involving 20 housekeeping genes (HKs), 14 genes encoding well-known cell envelope proteins (CEPs), 31 genes coding for secreted proteins (SECs), and 10 genes coding for proteins with unknown function or for which the biological role is not consensual. The SEC category involves proteins secreted [either by the Type III Secretion System (T3SS) - a machinery used by many bacterial pathogens to manipulate eukaryotic host cells by injecting virulence proteins - or by an undefined mechanism], into the cytosol of the host cells or to the inclusion membrane. For analyses enrolling divergence *versus* polymorphism, the corresponding orthologous genes of the *Chlamydia muridarum* Nigg strain were identified (by NCBI-BLAST search) and sequences were collected from the full-genome annotated in the GenBank database (accession number NC_002620) (29). For the strains that we needed to propagate as no *in silico* data was available, the 75 genes were amplified and sequenced by using standard procedures (111). The sequences and location of primers, as well as the amplicon sizes are listed in the Supplemental Table S2.3. Automated sequencing was achieved using BigDye Terminator v1.1 Cycle Sequencing chemistry, according to the manufacturer's instructions (Applied Biosystems) in an Applied Biosystems 3130*xl* Genetic Analyzer. Sequence reads were assembled using SeqBuilder software (DNASTAR) and alignments were generated using the ClustalW algorithm implemented in both the MegAlign software (DNASTAR) and MEGA5. A concatenated alignment of the 75 genes was also constructed for all *C. trachomatis* and *C. muridarum* strains. As the ClustalW program generates alignment artifacts in the presence of insertion/deletion (indel) events by disrupting codons, we edited "by hand" the amino acid alignments rather than only automate the process before editing the corresponding nucleotide sequences. When strain-exclusive single nucleotide polymorphisms (SNPs), indel events and pseudogenes were identified, resequencing was performed from a newly extracted DNA, and new sequences reads were generated for comparative purposes.

2.3.3. Phylogenomic analysis

Analyses of genetic diversity and phylogeny were conducted for each gene by using MEGA5. Briefly, we computed overall mean distances (number of differences and p-distance) and matrices of pairwise comparisons at both nucleotide and amino acid levels. For phylogenetic analysis, individual trees were generated using the Neighbor-Joining method with bootstrapping (234) and the

evolutionary distances were computed using the Kimura 2-parameter method (235). For all these analyses, the pairwise-deletion option was selected as it excludes sites containing alignment gaps or missing data from the analysis only when necessary in the pairwise distance estimation. Truncated genes, which are expected to encode non-functional proteins, were excluded from the phylogenetic and evolutionary analyses, except for the strains with non-disrupted sequences, as their biological role may be phenotype specific.

2.3.4. Global analysis of molecular evolution

The nonsynonymous/synonymous substitution rate ratio (d_N/d_S) among related protein-coding DNA sequences, where d_N refers to the number of nonsynonymous substitutions per nonsynonymous site and d_S is defined as the number of synonymous substitutions per synonymous site, may be suggestive of the selective pressures driving the mutational trends (236). Initially, for a global analysis of these trends, we estimated d_N and d_S values with MEGA5 by using the Kumar model (237). For each gene, d_N/d_S was calculated over all C. *trachomatis* sequence pairs and between the sequences of the two species (*C. trachomatis* and *C. muridarum*). More, in order to reinforce the comparison between the amount of evolutionary variation within the *C. trachomatis* species (polymorphism) and the variation between *C. trachomatis* and *C. muridarum* (divergence), we also applied the McDonald-Kreitman (MK) test (238,239). However, as it has been assumed that the results from the MK test cannot directly discriminate the type of selection acting on genes (240), the subjacent MK test algorithm was only used to clarify the neutral and amino acid-altering mutational trends underlying the *C. trachomatis* speciation process. This kind of analysis is suitable for tracing the Muller's ratchet phenomenon, which is commonly observed in niche-restricted pathogens.

2.3.5. Evaluation of the directionality in *C. trachomatis* evolution

In order to search for genes on which positive selection putatively operates, two distinct approaches were applied. First, as a statistical support of the d_N and d_S estimations within *C. trachomatis* strains, the codon-based Z-test of selection was computed by MEGA5 using the Kumar method (237), where bootstrapping (1000 replicates) was used for estimation of the variation in the statistic test. This test calculates the probability of rejecting the null hypothesis of strict-neutrality ($d_N = d_S$) in favor of one of two alternative hypothesis: positive selection ($d_N > d_S$) or purifying selection ($d_N < d_S$). Results with *P* values less than 0.05 were considered significant at the 5% level. On a second approach, the branch-site test of positive selection (branch-site test 2) (241,242) was employed using the *codeml* application from the Phylogenetic Analysis by Maximum Likelihood (PAML) package (version 4.4d) (243). Alignments of nucleotide sequences from the 59 *C. trachomatis* strains and *C. muridarum* (built and corrected on MEGA5) were converted into the "interleaved" PHYLIP format using the BioEdit package (version 7.0.0) (http://www.mbio.ncsu.edu/bioedit/bioedit.html), where stop codons were removed from sequences. The branch-site test is a robust bioinformatic approach (244) that is recommended to infer positive selection in a lineage of interest (called foreground lineage)

when several lineages in the phylogeny may have been subjected to distinct selective pressures (241,242). The statistical significance of the presence of positive selection along the branch of interest was addressed by the likelihood-ratio test (LRT) (245). In the branch-site test 2, the LRT compares the twice of the log likelihood difference (2Δl) between two models (alternative and null models) with the chi-square distribution with one degree of freedom for p-value calculation (242). The alternative model allows positive selection ($d_N/d_S \geq 1$) for the foreground branch, whereas the null model assumes the d_N/d_S ratios < 1 or = 1 for all site classes in all branches in the phylogeny. When positive selection acting on a specific gene was suggested by a significant LRT ($P < 0.05$), a Bayes empirical Bayes analysis (246) was used to identify the specific positively selected sites within that gene along the foreground branches. Therefore, the branch-site model requires an *a priori* definition and labeling of the foreground branches to be tested for positive selection, which should rely on well-defined biological hypotheses (241). Thus, based on the assumption that some genes might be involved in *C. trachomatis* phenotypic dissimilarities as a result of targeted positive selective pressures, we created six comprehensible biological hypotheses (H1-H6). The hypotheses evaluate the existence of genes under positive selection that may be involved in the following biological processes: specific cell-appetence to columnar epithelial cells of ocular (H1) or genital mucosae (H2), and to mononuclear phagocytes (H3); pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6). Only the genes for which the phylogeny supported any of these scientific hypotheses were tested.

Finally, as recombination may bias the results of positive selection, we used published data on recombination analysis enrolling all *C. trachomatis* genes (37,172,174) to inspect whether the genes selected for the present study showed evidences of recombination. Consequently, for the genes showing incongruent trees where unequivocal recombination was detected within a specific branch, the analysis of positive selection was excluded *a priori* for the corresponding biological hypothesis. On the other hand, genes yielding congruent trees but for which recombination had been previously detected (37,172,174) were still subjected to positive selection analysis and are properly identified in the present study.

2.3.6. Nucleotide sequences accession number

The nucleotide sequences determined in the present study were submitted to the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/index.html) and are currently available for consulting through the accession numbers: JQ066324 - JQ066722.

2.4. Results

2.4.1. Polymorphism significance of the selected genes

Distribution of point mutations in *C. trachomatis* chromosome is highly heterogeneous. Although the selected genes (Figure 2.1 and Supplemental Table S2.2) represent 11% of the coding

region length, they encompass about 55% of all chromosomal SNPs occurring within coding regions, which corresponds to a total of 5083 polymorphic sites among the 59 strains. In fact, we found that any given chromosomal SNP has 10.0 (odds ratio, 95% CI: 9.3 - 10.7) times higher probability to belong to the pool of genes under evaluation than to show up in any other gene (Fisher's exact test, $P < 10^{-7}$).

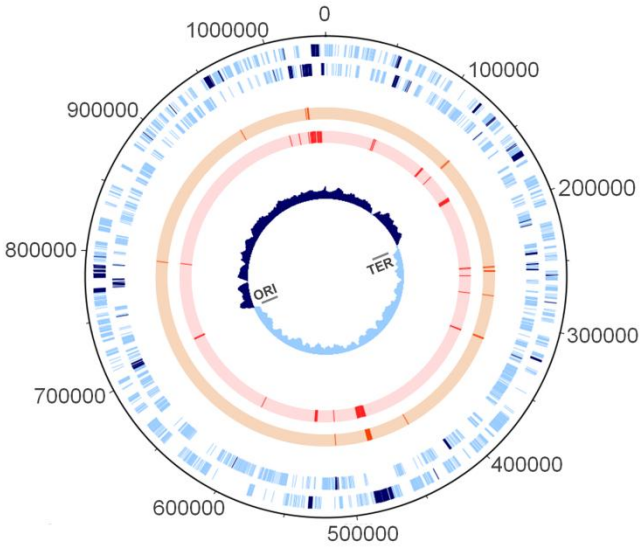


Figure 2.1. Chromosomal mapping of loci involved in the directional evolution of *Chlamydia trachomatis*. From the outside in, the first and second circles (light blue lines) refer to forward and reverse coding regions, respectively, according to the published genome of the *C. trachomatis* strain D/UW3. The 75 evolutionary informative genes evaluated in the present study are highlighted by dark blue lines. These loci encompass about 55% of SNPs occurring within the chromosomal coding regions. The third (orange lines) and forth circles (red lines) illustrate genes found to be under positive selection by the codon-based Z-test of positive selection (MEGA5) or the branch-site test of positive selection (PAML), respectively. Circle five shows the GC skew plot. The origin of replication (ORI) and the termination region (TER) are also marked. The figure was built using DNAPlotter (247).

2.4.2. Divergence *versus* polymorphism – detection of Muller's ratchet

Considering that *C. trachomatis* and *C. muridarum* species evolved from a last common ancestor (29), the comparison between the amount of evolutionary variation within the *C. trachomatis* species (polymorphism) and between *C. trachomatis* and *C. muridarum* (divergence) may shed some light on the evolutionary mutational dynamics that drove the *C. trachomatis* speciation. Accordingly, the divergence of the two species was evaluated through estimation of the d_N/d_S ratio between orthologous genes. All genes revealed d_N/d_S values lower than one, where the mean value was 0.21 (standard deviation [SD], 0.14) (Figure 2.2). This observation suggests an unequivocal higher weight of synonymous than nonsynonymous changes on the species divergence, in agreement with the

neutral theory of molecular evolution (248), which postulates that the fixation of selectively neutral mutations by random genetic drift is the major factor responsible for species divergence.

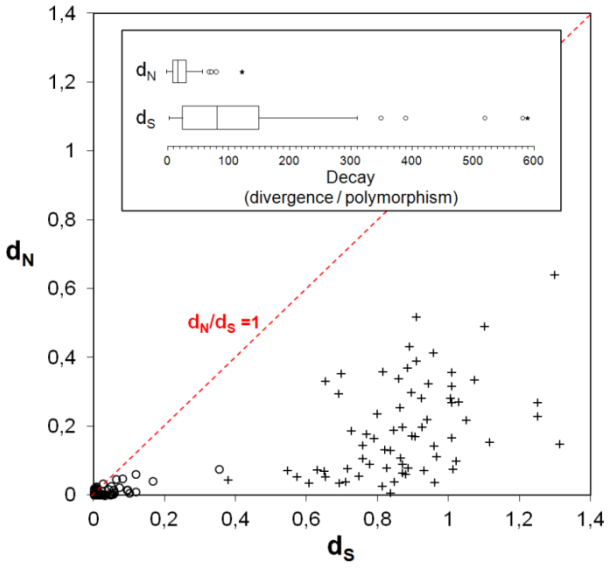


Figure 2.2. Evidence for Muller's ratchet phenomenon. These graphs show nonsynonymous *versus* synonymous mutational dynamics on the *C. trachomatis*/*C. muridarum* separation process. The scattering plot depicts the results concerning the evaluation of nonsynonymous and synonymous substitutions within the *C. trachomatis* species (open circles) (reflecting polymorphism) and between the species *C. trachomatis* and *C. muridarum* (crosses) (reflecting divergence). Neutrality line is also shown in red. The box plots display the dispersion of the overall decays in the d_N and d_S values [i.e., d_N(divergence)/d_N(polymorphism) and d_S(divergence)/d_S(polymorphism), respectively]. Outliers and extreme values are marked with open circles and asterisks, respectively. The considerable lower decay values for d_N suggests that the accumulation of deleterious mutations among strains from *C. trachomatis* species results from genetic bottleneck due to niche restriction (Muller's ratchet effect).

Subsequently, we compared these results with the d_N/d_S values obtained solely within *C. trachomatis* species (i.e., among the 59 strains). We observed that there was a high, but dissimilar, decrease of both d_S and d_N values after the separation of the two species from a common ancestor (Figure 2.2). In fact, the observed mean of the decay of d_S values [147.2 (SD±243.7)] was 5.5 times higher than the one observed for d_N [26.9 (SD±19.9)]. This was a consistent trend as the McDonald-Kreitman test algorithm yielded a similar decays ratio of 4.7. Globally, this suggests that, since *C. trachomatis* was established as species, the nonsynonymous changes increased their relative weight to synonymous changes in contrast with the divergence evolutionary process. This observation is consistent with the Muller´s ratchet theory (150,151) which assumes an accumulation of slightly deleterious non-silent mutations on microbial populations repeatedly subjected to genetic bottlenecks. In fact, the obligate intracellular lifestyle of *C. trachomatis*, which is characterized by niche-restricted and low-size populations, and expected low frequency of recombination relative to mutation events

(174,176) may lead to less effective elimination and consequent accumulation of non-silent mutations (153).

2.4.3. Evolutionary trends within the *C. trachomatis* species

We further focused on studying the accumulation of mutations after *C. trachomatis* speciation, with special emphasis on protein-altering changes that have contributed for phenotype divergences, which may help to clarify the *C. trachomatis* niche-specific adaptation. For each gene, the d_N/d_S ratio was evaluated over all C. *trachomatis* sequence pairs (Figure 2.3), where genes exhibiting d_N/d_S ratios above 1 and a significant *P* value (<0.05) in the codon-based Z-test of positive selection ($d_N/d_S > 1$) were considered as putative targets of positive selection. Twenty-seven genes exhibited overall d_N/d_S values higher than one, in which 15 (including 14 SECs) revealed a significant Z-test p-value (details in Supplemental Table S2.2). As predicted, all housekeeping genes presented d_N/d_S ratios below one, which indicates that the genes involved in regulatory/metabolic functions are less likely targeted by diversifying selection (249). On the other hand, 22 out of the 31 SECs support an opposite scenario, which is relevant as these proteins contact directly with the host, and thus are more prone to be involved in pathoadaptation.

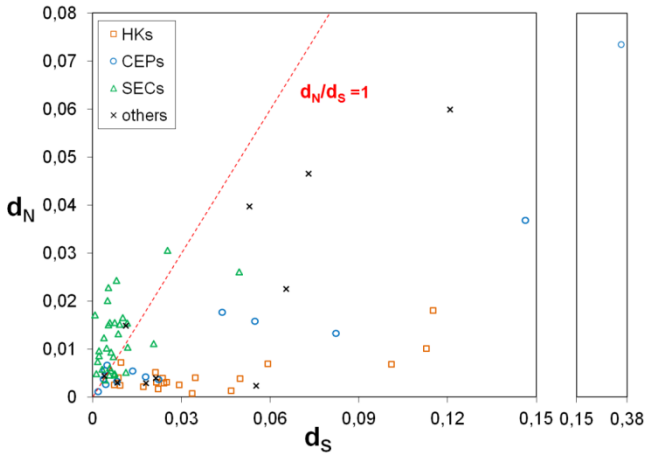


Figure 2.3. d_N *versus* d_S by gene functional category. This graph represents d_N and d_S values estimated for each gene for all 59 *C. trachomatis* strains. Housekeeping genes (HKs), genes encoding well-known cell envelope proteins (CEPs), genes coding for proteins secreted into the cytosol of the host cells or to the inclusion membrane (SEC), and "other genes" (see methods and Supplemental Table S2.2 for details) are represented by orange squares, blue circles, green triangles, and black crosses, respectively. Neutrality line is also shown in red.

We also investigated whether the types of mutation are dependent on the degree of genetic variability by evaluating the relationship between d_N and d_S values and the nucleotide polymorphism (p-distance) among *C. trachomatis* strains (Figure 2.4). Globally, as observed above for species divergence, we found that the increment in polymorphism is essentially driven by fixation of silent mutations, which presented an increase rate ~ 4-fold higher than for non-silent changes.

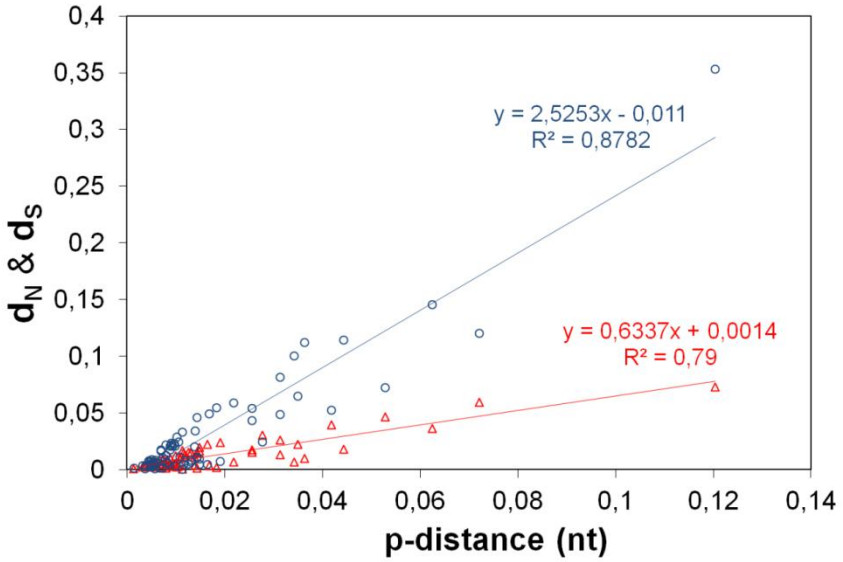


Figure 2.4. Genetic variability *versus* type of mutation. Distribution of d_N (red triangles) and d_S (blue circles) values according to the nucleotide polymorphism (mean genetic p-distance) of the 75 genes under evaluation among *C. trachomatis* strains. Slope values of the trend lines show a near 4-fold higher increase of d_S with p-distance than of d_N.

2.4.4. Distribution of d_N and d_S *versus* disease outcomes and ecological success

We also investigated whether the general distribution of both silent and non-silent mutations among *C. trachomatis* corresponds to strains clustering by disease outcomes. We used the concatenated sequences encompassing all genes under evaluation to calculate the d_N and d_S distances between each strain and the different groups of strains (i.e., three disease groups) (Figure 2.5). Our results sustain a non-random accumulation of mutations where strains with the same cell-appetence are unequivocally clustered either by silent mutations or protein-changing alterations. Indeed, the genetic distances between strains with dissimilar tropism are 1.8- to 10.5-fold (for d_S) and 1.7- to 6.7-fold (for d_N) higher than the ones between strains with similar cell-appetence. Additionally, the highly ecological succeeded strains causing non-invasive genital infections (mostly from serovars E and F) are slightly separated from the remainder genital strains (Figure 2.5). This analysis clearly supports that our approach for detecting positively selected genes (and codons) relying on rationally established biological hypotheses (see methods) may be an useful step for understanding the molecular basis underlying *C. trachomatis* phenotypic differences.

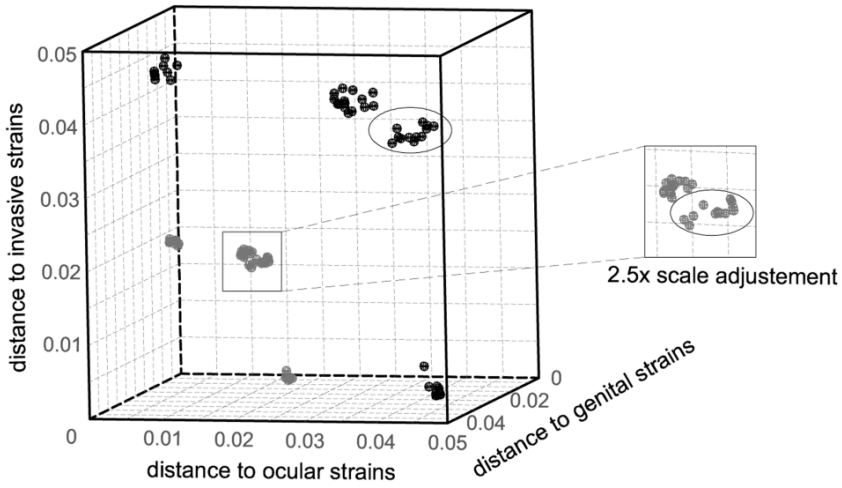


Figure 2.5. Non-random distribution of both nonsynonymous and synonymous mutations according to tropism and ecological success. The 3D scatter plot shows the genetic distances between each of the 59 strains and the three disease groups by both d_N (grey) and d_S (black) estimations. Values were estimated by using the concatenated sequences enclosing all genes under evaluation. Strains infecting mononuclear phagocytes (shown in the bottom of the cube), the columnar epithelial cells of ocular (shown in the left side) or genital mucosae compose three major clusters for both d_S and d_N. Within the non-invasive genital strains, the more clinically prevalent strains (labeled with an ellipse) are clustered apart.

2.4.5. Positive selection driving bacterial specific appetence to different human cell types

The phylogenetic analysis revealed genes whose trees cluster all strains that preferentially infect the same human cell type in a single branch. Thus, genes (and codons) targeted by positive selection along those branches may be involved on specific host-cell interactions. To evaluate this, we conducted the branch-site test of positive selection under the biological hypotheses H1 to H3 (see Material and Methods). All genes and the inferred positively selected codons found to be putatively involved in specific adaptive processes are described in Figure 2.6 and Table 2.1.

Five genes were found to be under positive selection in the evolutionary process that drove the segregation of ocular strains (branch H1). These include genes encoding two polymorphic membrane proteins (Pmps) (CT869/*pmpE* and CT870/*pmpF*), one Pmp-like protein (CT050), one inclusion membrane protein (Inc) (CT115/*incD*) and the translocated actin recruiting phosphoprotein (CT456/*tarp*). The Pmps and Incs are among the most promising research targets for which there is cumulative evidence of their involvement in biological mechanisms such as adhesion, immune response elicitation, or subversion of intracellular trafficking (see Table 2.1 for details) (112,250-255). For example, PmpF was predicted *in silico* to contain T-cell epitopes that bind HLA class I and II alleles (165). On the other hand, Tarp is a chlamydial effector of the T3SS associated with the

chlamydial invasion of the host cells(141) by mediating host actin polymerization and inclusion development (140,256).

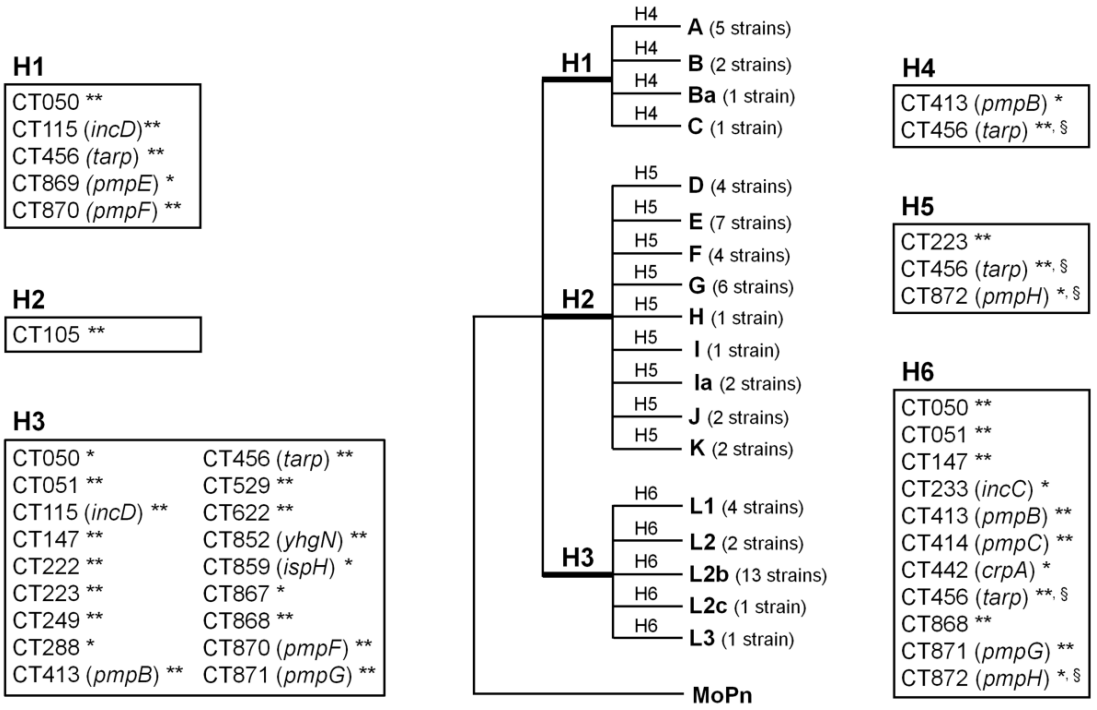


Figure 2.6. Positive selection driving the directional *C. trachomatis* evolution towards niche-specific adaptation. The figure represents a model tree encompassing all 59 *C. trachomatis* strains that was created to facilitate a proper visualization of all biological hypotheses. These evaluate the existence of genes under positive selection (through the branch-site test of positive selection) that may be involved in: specific cell-appetence to columnar epithelial cells of ocular (H1) (serovars A-C) or genital mucosae (H2) (serovars D-K), and to mononuclear phagocytes (H3) (serovars L1-L3); pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6). This test was applied to each individual gene tree in an independent manner for each hypothesis depending of the tree topology. The figure boxes show the genes found to be positively selected for each biological hypothesis. The likelihood-ratio test (LRT) was used for inferring the statistic significance (p-values) of positive selection in the foreground branches. ** and * indicate significance with $P < 0.01$ and $P < 0.05$, respectively. [§] refers to genes presenting congruent trees, but for which a specific biological hypothesis may be affected by recombination (174). See Table 2.1 for details on positively selected codons.

Table 2.1. Positively selected genes and the inferred codons putatively involved in specific adaptive evolution based on the branch-site test of positive selection by PAML.

ORF [a] (gene)	Biological hypothesis with positive selection (H1 to H6) [b]	p-value (LRT test) [c]	Specific codons under positive selection [d, e]	Putative function / experimental evidences	Specific biological hypothesis excluded due to putative recombination [f]
CT050	H1	$p < 10^{-10}$	ND	Pmp-like protein identified in the inclusion lumen (156,257).	H2, H5
	H3	0.0352	51K*, 80K*, 366K*, 523S*, 546K*		
	H6	$p < 10^{-7}$	183N*, 184S*, 223D**		
CT051	H3	0.0004	563V*	Pmp-like protein (156,257)	H2, H5
	H6	0.0002	435R**		
CT105	H2	0.0004	ND	Function unknown.	---
CT115 (*incD*)	H1	$p < 10^{-4}$	11D**, 12G*	Inc (253). T3SS effector (132)	H2, H5
	H3	$p < 10^{-9}$	150L*, 151E*, 153S*, 155S**		
CT147	H3	$p < 10^{-5}$	1407I*	Inc; human early endosomal antigen 1 (EEA1) homologue (99) Immunodominant antigen (157). Involvement in pathogenic differences *in vivo* (164).	H1, H2, H4, H5
	H6	0.0003	ND		
CT222	H3	$p < 10^{-7}$	124I*,125S**,126V**	Inc associated with host kinases in microdomains that interact with the host centrosomes (258).	---
CT223	H3	0.0051	127G*, 206R*	Inc (259). T3SS effector (132). Subversion of intracellular trafficking (255). Host cell cytokinesis blockage (260).	---
	H5	0.0082	99R*, 152S*		
CT233 (*incC*)	H6	0.0432	5M*,6S**,7D**,8I*, 11K*, 14I*	Inc (261). T3SS effector (130,132).	---
CT249	H3	0.0004	8Y*, 24N*, 80T**, 89I**	Inc (262). T3SS effector (129).	---
CT288	H3	0.0104	ND	Inc (259). T3SS effector (132).	H1, H2, H4, H5
CT413 (*pmpB*)	H3	0.0015	ND	Adhesin (112). Antigen (250-252).	H2, H5
	H4	0.0198	58A*, 75T**, 235A*, 820A*, 947E*, 998A*, 1061N**, 1171T**		
	H6	0.0017	ND		
CT414 (*pmpC*)	H6	$p < 10^{-6}$	145P*, 544I**,598V**	Adhesin (112). Antigen (250-252).	H2, H5
CT442 (*crpA*)	H6	0.0154	24A*, 29K**, 48I*, 104I**, 133D*, 137V**	Inc (259).T3SS effector (132). Antigen (159).	---
CT456 (*tarp*)	H1	0.0094	447H*, 978H*	Translocated actin-recruiting phosphoprotein / early T3SS effector (140). Contribution for the pathogen phagocytosis (141). Involvement in pathogenic differences *in vivo* (164). Antigen (263,264).	---
	H3	$p < 10^{-24}$	ND		
	H4	0.0049	ND		
	H5	$p < 10^{-8}$	189A**, 237G*, 407S*, 481A**		
	H6	$p < 10^{-28}$	134N*, 139I**, 252G**, 301D*, 351D**, 399G*,		

			404D**, 493D**[,§], 494D*[,§], 530K*, 577R*, 603A*, 888G*, 891D**, 909K*		
CT529	H3	0.0009	3A*	Inc (160). T3SS effector (132).	---
CT622	H3	0.0012	ND	Antigen (160,263). Involvement in pathogenic differences *in vivo* (164). T3SS effector (265). Antigen (252,264).	H1, H2, H4, H5
CT694	H? [g]	0.0190	ND	Immunodominant antigen (157,252). Early T3SS effector (266). Modulation of host cell processes (267).	---
CT852 (*yhgN*)	H3	$p < 10^{-4}$	202M**, 203L*	Putative integral membrane protein (46).	H2, H5
CT859 (*ispH*)	H3	0.0135	ND	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (115)	H2, H5
CT867	H3	0.0141	ND	Deubiquitinase and deneddylase (268).	H1, H2, H4, H5
CT868	H3	0.0064	ND	Deubiquitinase and deneddylase (268).	H1, H2, H4, H5
	H6	$p < 10^{-5}$	9S**, 17R**, 26S**, 30R**, 105T**[,§], 136D**, 138L**, 157Q**, 158T**, 222P**, 238R**, 244V**, 256S**, 287R**, 300P**, 303N**, 307E**, 311F**, 322Y**, 323D**, 324S**, 325K**, 339R**, 340G**, 343S**, 354H**, 358K**, 361L**	Involvement in pathogenic differences *in vivo* (164). Inhibition of a crucial pathway for host inflammatory responses (269).	
CT869 (*pmpE*)	H1	0.0348	59N*, 139I*, 469A*	Adhesin (112). Antigen (250,251).	H2, H5
CT870 (*pmpF*)	H1	$p < 10^{-4}$	ND	Adhesin (112). Antigen (165).	---
	H3	0.0020	ND		
CT871 (*pmpG*)	H3	0.0082	ND	Adhesin(112) Antigen (250-252).	H2, H5
	H6	$p < 10^{-4}$	258G*, 320K*, 324S*, 812Q**		
CT872 (*pmpH*)	H5	0.0303	ND	Adhesin (112). Antigen (250,252).	---
	H6	0.0302	ND		

[a] Open reading frame (ORF) numbers are based on genome annotation of the strain D/UW3 (GenBank No. NC_000117).
[b] The hypotheses were created to evaluate the existence of genes under positive selection involved particular biological processes: specific cell-appetence to columnar epithelial cells of ocular (H1) or genital mucosae (H2), and to mononuclear phagocytes (H3); pathogenic diversity among strains causing ocular disease (H4), genital disease (H5), or hemorrhagic proctitis and suppurative lymphadenitis (H6).
[c] The likelihood-ratio test (LRT) was used to infer the statistic significance of positive selection in the foreground branches (p-value). The degree of freedom is 1 for the comparisons of alternative hypothesis *versus* the null hypothesis in the branch-site test 2.
[d] The posterior probabilities that each site belongs to the site class of positive selection on the foreground lineages are inferred by the Bayes Empirical Bayes (BEB) analysis. Positively selected sites are those with ** $P > 0.99$ and * $P > 0.95$. ND, not discriminated (an excess of positively selected codons hampered their discrimination by PAML, or the identified codons revealed $P < 0.95$).
[e] For simplification purposes, amino acid positions for biological hypotheses H1 and H4 are based on the protein sequence annotation for the strain A/Har13, whereas for H2 and H5 are based on the annotation for strain D/UW3, and those for H3 and H6 refer to strain L2/434. Within CT456/*tarp* and CT868, the position of the codons labeled with [§] symbol refer to strain L1/1322/p2 since L2/434 is deleted in this region.
[f] Recombination was detected in previous studies (37,172,174). ---, no hypothesis excluded.
[g] This gene was detected to be under positive selection specifically for the most clinically prevalent genital strains (mostly from E and F serovars).

For the biological hypothesis H2, we found significant evidence supporting a fixation of adaptive mutations driving a better appetence to the columnar epithelial cells of genital mucosa (branch H2) solely for the gene CT105. Although its function is unknown, a previous study of heterologous expression in yeast (156) suggested that CT105 may be involved in modulation of host cellular functions. Nevertheless, it is worth to note that CT105 is a pseudogene for ocular strains, thus we may be facing a scenario of a gene strictly needed for tropism functions other than those involving the ocular conjunctiva. The frequent tree incongruence involving genital strains hampered the evaluation of several genes for this specific biological hypothesis (Table 2.1).

Regarding the branch that clusters all strains that infect the mononuclear phagocytes (branch H3), we detected 18 genes likely under positive selection. This set includes one HK, and genes encoding 11 SECs (seven Incs), four CEPs (including three Pmps), and two proteins with unknown function (Table 2.1). Besides the general relevance of Incs and Pmps (explained above), we would highlight the SEC CT223, an Inc protein for which it was suggested a role in subversion of host cell functions, either by containing SNARE-like (eukaryotic soluble N-ethylmaleimide-sensitive attachment protein receptors) motifs (impact in intracellular trafficking) (255), or by blocking host cell cytokinesis (260). The SECs CT622 is an antigen putatively secreted by the T3SS (252,264,265), whereas CT867 and CT868 are proteases that possess deubiquitinating and deneddylating activities (268), which may suggest a role in virulence. The invasive infection pattern of L1-L3 strains along with the expectation that these strains were the first to diverge from a common *C. trachomatis* ancestor (270) may justify the high number of genes detected under diversifying selection on branch H3 rather than on branches H1 and H2. These genes may play a role in the specificity of mononuclear phagocytes-bacteria interactions that yield invasive infections with L1-L3 strains.

2.4.6. Positive selection driving pathogenic diversity among strains infecting the same human cell type

The detection of positive selection acting on specific genes along branches of strains causing similar disease outcomes may be useful for understanding adaptive alterations underlying niche-specific pathogenic dissimilarities. Among the ocular strains, we detected two genes under positive selection (branches H4) (Figure 2.6): one *pmp* (CT413/*pmpB*) and CT456/*tarp* (Table 2.1). CT456/*tarp* had been already indicated as potentially involved in distinct pathogenic patterns displayed by two ocular strains (both of serovar A) when infecting cynomolgus monkeys (164).

Regarding the detection of positive selection driving pathogenic differences among strains causing non-invasive urogenital disease (branches H5), we found the above described virulence factors CT223 and CT456/*tarp*, and one *pmp* (CT872/*pmpH*). Once more, the analysis of positive selection underlying pathogenicity among non-invasive genital strains was impaired by the described recombination events involving these strains (Table 2.1) (37,172,174). Of notice was the detection of positive selection events governing the evolutionary segregation of the most succeeded genital strains (mostly from E and F serovars) for the gene coding CT694. This protein is an immunodominant antigen (157), and it was also demonstrated its secretion into the host cytoplasm by the T3SS at early

time-points after infection (as CT456/Tarp), where it localizes to host cell membranes and interacts with eukaryotic AHNAK, an actin-binding protein; it is believed that CT694 may act by regulating membrane fluidity or by remodeling actin filaments during invasion or early stages of *C. trachomatis* development (266,267). Thus, differences in immune evasion strategies or in host-cell manipulation during invasion may be crucial biological processes underlying ecological success.

Despite the remarkable genomic homogeneity of *C. trachomatis* strains that infect the mononuclear phagocytes (Figure 2.5), we found 11 positively selected genes along the branches embodying SNPs that distinguish these strains (branches H6). These include four *pmps* (CT413/*pmpB*, CT414/*pmpC*, CT871/*pmpG* and CT872/*pmpH*), two genes encoding Pmp-like proteins (CT050 and CT051), three genes coding for Incs (CT147, CT233/*incC* and CT442/*crpA*), CT456/*tarp* and CT868. Besides what was generally described above for these proteins, it is worth to note that the Inc CrpA (259) is a T3SS substrate (132) that may play a role in immune evasion as it was found to be targeted by CD8+ T cells in response to infection in murine (159). Also notable is that all adaptive codons inferred for the CT233/*incC* (Table 2.1) correspond to the IncC N-terminal domain, and, more specifically within the first 15 residues, where it is known to reside the secretion signals recognized by T3SSs (132).

2.4.7. Evolutionary inferences and associated bias

The branch-site test may generate some erroneous detected positively selected genes. Although this was not assessed in the present study, recent robust evaluations estimated a range of 0-5% of false positives on this test (244). Nevertheless, this is more problematic when performing inter-species analyses as it involves highly divergent sequences (271), which is not the case of the intra-species analysis performed in the present study. Furthermore, to guide against violations of model assumptions, we applied very conservative criteria to calculate p-values in the LRT by using χ_1^2 as the null distribution (242,271). Recombination is another critical factor that may bias the estimation of positive selection. As previous data based on full-genome sequences (37,172,174) detected recombination for some of the genes enrolled in the present study, some specific biological hypothesis could not be subjected to the branch-site test of positive selection. These specific exclusions are indicated in the Table 2.1. The remainder biological hypotheses were tested as recombination is not observed in the corresponding branches. For example, for CT147 tree, where some ocular strains are shown within a genital branch (hampering the analysis of the hypotheses H1, H2, H4, and H5), the hypotheses H3 and H6 could still be validated. In another scenario, when recombination is known to occur in genes presenting strong congruent trees (as for CT870/*pmpF*), all hypothesis were evaluated. For these specific cases (indicated in Figure 2.6) the results should be eyed with caution.

2.5. Discussion

A well-known metaphor in evolutionary biology is the adaptive landscape represented by a two-dimensional plot of all genotypes in a specific environment, with their fitness represented by the

height of the landscape. For each new environment, in order to climb the fitness peak, bacteria will have to acquire new beneficial mutations, which will likely be differentially spread among different genotypes (272). Presumably, the radiation of *C. trachomatis* species into strains with different cell-appetence may be explained by this scenario. Indeed, the different environments are represented by the dissimilar human tissues that strains preferentially infect (ocular, genital tract and lymph nodes), which present heterogeneous properties in terms of competing flora, immune response, and physiological characteristics (such as pH and hormonal concentration). Also, strains present dissimilar fitness as serovar E and F together represent more than 40% of all genital infections worldwide (63), and serovars A and L2 clearly predominate in ocular (53) and lymphogranuloma (46) infections, respectively. On the other hand, *C. trachomatis* is an obligate intracellular bacterium with low doubling time (231) and population size (273), and is thus subjected to transmission bottlenecks which make this pathogen a target for the accumulation of deleterious mutations on behalf of the Muller's ratchet theory. The validity of the Muller's ratchet has been evaluated either in RNA viruses (274), which present high mutation rates, are subjected to recurrent bottlenecks and the rate of compensatory back-mutations is low, or even in large free-living bacteria such as *Salmonella typhimirium*, where these contributing factors are clearly attenuated (153). In our study when the values of d_N and d_S are compared independently, it is noticeable that, after the *C. muridarum/C. trachomatis* separation from a common ancestor, the values of d_S show a 147-fold decrease whereas the values of d_N only decrease 27-fold (Figure 2.2), which suggests the existence of an accumulation of deleterious mutations due to genetic bottleneck, as postulate by Muller. We speculate that, in addition to this scenario, some non-silent changes may reflect adaptive mutations to different niches rather than deleterious mutations specific of same-niche infecting strains. Thus, the detection of positive selection events acting on particular genomic regions may help to elucidate genotype-phenotype relationships. Unfortunately, there are scarce cases where genotype and phenotype are unequivocally linked in *C. trachomatis*, because no straightforward tools to genetically manipulate this pathogen are available so far. One of the few examples is illustrated by the mutational pattern of the *trpBA* operon (encodes tryptophan synthase, which uses indole as substrate), where genital strains possess an intact and active operon whereas it is truncated by point mutations or small indels in strains infecting the ocular conjunctiva (where indole is rare) (51). Thus, mutations that are beneficial in one genetic background are not necessarily beneficial in another background. In our study, a similar scenario may stand for CT105, which is a pseudogene for the ocular strains, whereas our results suggest that it may be involved in the strains' appetence to the genital epithelium (Figure 2.6 and Table 2.1).

Our results showed that non-silent changes differentiate strains with different cell-appetence or pathogenesis (Figure 2.5) and involve genes whose functions may underlie distinct phenotypes. In fact, among the 25 genes identified as positively selected along specific lineages (Table 2.1), we found genes encoding proteins implicated in immune response elicitation (such as CT147, CT442/*crpA,* CT529, CT694, and *pmps*) (157,159,160,250,252,264), proteolytic activity (such as CT867 and CT868) (268), and subversion of host-cell functions (such as CT223 and CT456/*tarp*) (141,255). Some of these genes were also identified in a previous study (174), but no genotype/phenotype associations could be established because only six serovars were evaluated,

contrarily to the present study which constitutes a considerable scale-up in terms of genetic variability (enrolling all major 15 serovars represented by 59 strains). A detailed view of the positively selected *loci* that we have detected revealed 11 genes (CT050, CT051, CT115/*incD*, CT147, CT223, CT413/*pmpB*, CT456/*tarp*, CT868, CT870/*pmpF*, CT871/*pmpG* and CT872/*pmpH*) supporting two or more biological hypotheses for adaptive changes (Figure 2.6). Although this seems intriguing in terms of evolutionary directionality, experimental evidences suggest multiple functions for some of them (Table 2.1) or a distinct *modus operandi* for a specific function. The most striking example is illustrated by CT456 which codes for Tarp. Strong experimental evidences showed that this T3SS effector is associated with the recruitment of host-cell actin observed at early stages of invasion, involving a C-terminal actin binding domain (ABD) and a proline-rich region (275). Whereas the invasive serovar L2 contains a single functional ABD and it is believed that the proline rich domain plays also a role in actin nucleation, Tarp from strains with different cell-appetence contain multiple ABD sites that are able to nucleate actin without the need of the respective proline-rich domain (141). These data suggest that strains may use Tarp distinctly for actin nucleation. Also, Tarp harbors an N-terminal tyrosine-rich repeat domain (the number of repeats are serovar-dependent) that is tyrosine phosphorylated by host cell kinases (256). Curiously, some positively selected sites found to be associated with infection of mononuclear phagocytes (Table 2.1) are located precisely within the tyrosine-rich repeat domain. Moreover, there seems to be a pattern of amino acid substitution, where positive selection is operating on exactly the same amino acid positions within the repeated regions, involving always the exchange between aspartic acid (D) and glycine (G) for seven of the positively selected sites (Table 2.1). In support of recently published data by Mehlitz and colleagues (276), it can be speculated that, as for the Tarp C-terminal region, dissimilar *modus operandi* of Tarp N-terminal region may underlie distinct phenotype properties of *C. trachomatis* strains. Also of relevance is CT868, a deubiquitinating and deneddylating enzyme that likely interferes in multiple cellular processes. Indeed, a recent study demonstrated that CT868 is capable of inhibiting the host inflammatory responses by blocking the nuclear factor-kB pathway, a known mechanism by which pathogenic microorganisms evade the host immune responses (269). Facing these data and considering the privileged representation of these 11 *loci* in the Figure 2.6, they should be considered as crucial research targets to improve our knowledge in the pathobiology of *C. trachomatis*.

It is also worth notable that, among the nine-member Pmp paralogues family, six genes seem to be under positive selection for specific phenotypes. Cumulative evidences indicate that Pmps may function as fine-tune determinants of *C. trachomatis* pathobiology, either by antigenic variation (165,250,252,277), or host-cell adhesion (110-112), hypothetically through a shut-off mechanism at the inclusion level (278). It is posited that the accelerated evolution between paralogues is common and constitutes a mechanism for the generation of new genes and new biochemical functions (279).

Darwinian evolution generally relies on the existence of an adaptive pathway in which intermediate steps provide a gradual improvement of fitness. Thus, adaptive changes should not completely rule out synonymous mutations. In fact, the latter may alter the immediate protein adaptive landscape (by changing the proximal amino acids), providing the protein with new opportunities to evolve (280). Also, they can change RNA secondary structure and influence its stability (281) as well

as originate codons with different frequency usage (associated with tRNA abundance), which was already shown to affect the translation efficiency in several microorganisms (282,283). We have previously shown synonymous changes to more favorable codons for the *C. trachomatis* major antigen (63), and it is reasonable to expect that several other *loci* present synonymous changes that, in a camouflaged way, become adaptive. Our results support a non-random accumulation of synonymous mutations in *C. trachomatis*. In fact, we found that strains infecting the same human cell type are clearly the most closely related through d_S analysis (Figure 2.5), which suggests a nonstochastic fixation of synonymous mutations.

In conclusion, our results support a directional evolution of *C. trachomatis* towards niche-specific adaptation in addition to a background of Muller's ratchet deleterious mutations. Whereas the molecular basis for organ/cell-appetence is likely complex, these data suggest that population genetics and evolutionary inferences may be crucial to a comprehensive understanding of the resulting phenotypes, by guiding subsequent experimental procedures to specific targets.

Acknowledgements

This work was supported by a grant, PTDC/SAU-MII/099623/2008, from Fundação para a Ciência e a Tecnologia (FCT). VB and RF were recipients of Ph.D. fellowships (SFRH/BD/68527/2010 and SFRH/BD/68532/2010, respectively) from FCT. AN was a recipient of a post-doctoral fellowship (SFRH/BPD/75295/2010) from FCT.

Polymorphisms in Inc proteins and differential expression of *inc* genes among *Chlamydia trachomatis* strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates

This chapter corresponds to the contents (with minor changes) of the following publication. The decision to include contents that were not performed by the author of the present Ph.D. thesis relies on the need to put the data into the appropriate context.

2012, Journal of Bacteriology, 194:6574-6585.
Almeida F.*, Borges V.*, Ferreira R., Borrego M.J., Gomes J.P., Mota L.J.
Polymorphisms in Inc proteins and differential expression of *inc* genes among *Chlamydia trachomatis* strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates.
(http://www.ncbi.nlm.nih.gov/pubmed/23042990)

* These two authors equally contributed to this work.

Personal contribution

VB contributed to the design of the study, co-performed culture, RT-qPCR assays, analyses of genetic variability, co-interpreted data and revised the manuscript.

3. Polymorphisms in Inc proteins and differential expression of *inc* genes among *Chlamydia trachomatis* strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates

3.1. Abstract

Chlamydia trachomatis is a human bacterial pathogen that multiplies only within an intracellular membrane-bound vacuole, the inclusion. *C. trachomatis* includes ocular and urogenital strains, usually causing infections restricted to epithelial cells of the conjunctiva and genital mucosa, respectively, and lymphogranuloma venereum (LGV) strains, which can infect macrophages and spread into lymph nodes. However, *C. trachomatis* genomes display > 98% of identity at DNA level. In this work, we studied whether *C. trachomatis* Inc proteins, which have a bilobed hydrophobic domain that may mediate their insertion in the inclusion membrane, could be a factor determining these different types of infection and tropisms. Analyses of polymorphisms and phylogeny of 48 Inc proteins from 51 strains encompassing the three disease groups showed significant amino acid differences that were mainly due to variations between Inc proteins from LGV and ocular or urogenital isolates. Studies of the evolutionary dynamics of *inc* genes suggested that 10 of them are likely under positive selection and indicated that most non-silent mutations are LGV specific. Additionally, real-time quantitative PCR analyses in prototype and clinical strains covering the three disease groups identified three *inc* genes with LGV-specific expression. We determined the transcriptional start sites of these genes and found LGV-specific nucleotides within their promoters. Thus, subtle variations in the amino acids of a subset of Inc proteins and in the expression of *inc* genes may contribute for the unique tropism and invasiveness of *C. trachomatis* LGV strains.

3.2. Introduction

Chlamydiae are a large group of obligate intracellular bacteria. They include *Chlamydia trachomatis*, which causes ocular and genital infections in humans. These infections are the leading cause of preventable blindness in developing countries (6), and the most prevalent cause of bacterially sexually transmitted diseases worldwide (4). *C. trachomatis* strains include trachoma and LGV (lymphogranuloma venereum) biovars (50). The trachoma biovar comprises ocular and urogenital strains, which cause localized infections of the epithelial surface of the conjunctiva and genital mucosa, respectively; strains of the LGV biovar cause invasive urogenital disease, due to their ability to infect macrophages and spread into lymph nodes. *C. trachomatis* strains can be further classified into ocular serovars A-C; urogenital serovars D-K; and LGV serovars L1-L3.

C. trachomatis genomic sequences of different ocular, urogenital and LGV strains exhibit >98% of identity and a high degree of synteny (37,46,115,164-166,172,175,284). Therefore, the determinants of the different types of infection (invasive or non-invasive) and tissue tropism (eyes, genitals, and lymph nodes) must rely on the few genes present in some strains but not in others, and on nucleotide differences which may either lead to proteins with disease group-specific amino acids or

to differential gene expression. Some of these determinants have been suggested in previous studies: the tryptophan (*trpRBA*) operon (51,180,285) and genes encoding cytotoxin (179), phospholipase (286), polymorphic membrane proteins (Pmps) (111), and Tarp (287).

Chlamydiae are characterized by a developmental cycle involving the interconversion between an infectious form, the elementary body, and a non-infectious form, the reticulate body (88). Throughout development, the bacteria reside and multiply within a membranaceous compartment, known as the inclusion, and manipulate host cells by using a type III secretion (T3S) system to translocate effector proteins into host cells (103,127). Chlamydial T3S substrates have been found by the identification of an N-terminal secretion signal using *Salmonella* (288), *Shigella* (132,289), or *Yersinia* (130,140,266,290,291) as heterologous hosts. These include inclusion membrane (Inc) proteins, characterized by a bilobed hydrophobic motif thought to mediate their insertion into the inclusion membrane (129-132,289). Incs from the same chlamydial species are normally unrelated to each other (129,292), and only a subset of ~25 Inc proteins is conserved between species (129,292). *C. trachomatis* Inc proteins CT119/IncA, CT115/IncD, CT147, CT229, and CT813 have been shown or suggested to subvert host cell vesicular and non-vesicular transport (255,293,294). However, virtually nothing is known about the biological role of most Inc proteins, which also reflects the lack of straightforward methods to genetically manipulate *Chlamydiae*.

In this work, we used phylogenetic, molecular evolution, and gene expression analyses to analyze whether Inc proteins could affect the type of infection and tissue tropism associated with *C. trachomatis*. Our studies suggest that a subset of Inc proteins might play a role in the unique capacity of *C. trachomatis* LGV strains to infect macrophages and disseminate into lymph nodes.

3.3. Materials and Methods

3.3.1. Bacterial strains and growth conditions

C. trachomatis prototype strains B/Har36, C/TW3, E/Bour, L2/434, and L3/404 (from ATCC) and clinical strains F/CS465-95 and L2b/CS19-08 (from the collection of the Portuguese National Institute of Health) were used as detailed below, and propagated in HeLa 229 cells (from ATCC) using standard techniques (193).

Escherichia coli TOP10 (Invitrogen) was used for construction and purification of the plasmids. *Yersinia enterocolitica* ΔHOPEMT (MRS40 pIML421 [$yopH_{\Delta 1\text{-}352}$, $yopO_{\Delta 65\text{-}558}$, $yopP_{23}$, $yopE_{21}$, $yopM_{23}$, $yopT_{135}$]) (295), deficient for the *Yersinia* T3S effectors YopH, O, P, E, M, and T, but T3S-proficient, was used for T3S assays. To construct a T3S-deficient derivative of ΔHOPEMT, we deleted in this strain the complete coding sequence (codons 1-354) of the *yscU* gene, which encodes an essential component of the *Y. enterocolitica* T3S system (296). This was done by allelic exchange with the mutator plasmid pLY16 (296). The resulting *Y. enterocolitica* ΔHOPEMT ΔYscU strain was also used in T3S assays. The $yscU_{\Delta 1\text{-}354}$ mutation had been previously shown to be non-polar (296). *E. coli* or *Y. enterocolitica* were routinely grown in liquid or solid Luria-Bertani medium with the appropriate

antibiotics and supplements. Plasmids were introduced into *E. coli* or *Y. enterocolitica* by electroporation.

3.3.2. Construction of plasmids

Plasmids were constructed and purified using proof-reading Phusion DNA polymerase (Finnzymes), restriction enzymes (MBI Fermentas), T4 DNA Ligase (Invitrogen), DreamTaq DNA polymerase (MBI Fermentas), NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), and GeneElute Plasmid Miniprep kit (Sigma), according to the instructions of the manufacturers. In brief, to analyze T3S signals we constructed plasmids harboring hybrid genes encoding the N-terminal 20 amino acids of each *C. trachomatis* Inc, or of the *Y. enterocolitica* T3S chaperone SycT (295), and the mature form of TEM-1 β-lactamase (TEM-1). These hybrids were made by PCR using plasmid pCX340 as template (297). Each forward primer contained an *Nde*I site followed by 60 nucleotides encoding the first 20 amino acids of each Inc (or of SycT) in frame with a sequence complementary to the 3' extremity of the transcribed strand of the TEM-1-encoding gene; each reverse primer contained either a *Hin*dIII or a *Xho*I site followed by a sequence complementary to the 3' extremity of the non-transcribed strand of the TEM-1-encoding gene. Digested PCR products were ligated into pLJM3, a low-copy plasmid which enables expression of cloned genes driven by the promoter of the *Y. enterocolitica yopE* gene (298). A similar strategy was used to construct a plasmid encoding TEM-1 alone, except that the forward primer did not contain an *inc* gene sequence. For primer design, the DNA sequence of each *inc* gene in *C. trachomatis* strain L2/434 (see Supplemental Table S3.1) was used, except for *ct036*, *ct115/incD*, and *ct119/incA*, in which the sequence from strain D/UW3 was used (see Supplemental Table S3.1); the DNA sequence of *sycT* was from *Y. enterocolitica* pYVe227 (accession number AF102990). The sequence of all primers is available upon request. The accuracy of the nucleotide sequence of all the inserts in the constructed plasmids was checked by DNA sequencing.

3.3.3. *Y. enterocolitica* T3S assays

These analyses were done as previously described (296). Proteins in bacterial pellets and culture supernatants were analyzed by immunoblotting with mouse monoclonal anti-TEM-1 antibodies (QED Bioscience; 1:500), and with rabbit polyclonal anti-SycO antibodies (1:1000) (299), and detected with a ChemiDoc XRS+ system (BioRad). The amount of protein in the supernatant relative to the amount of total protein (% of secretion), was determined from immunoblot images with Image Lab (Bio-Rad).

3.3.4. DNA sequences

Regardless of the origin of the *C. trachomatis* gene, throughout this work we used the nomenclature of the annotated D/UW3 strain (see Supplemental Table S3.1). The nucleotide

sequences of the *inc, pmp*, and housekeeping genes analyzed were from the available genomes of 51 *C. trachomatis* strains (37,46,115,164-166,172,175,284). The strains and corresponding genome accession numbers are listed in see Supplemental Table S3.1. The DNA sequences were retrieved from pairwise alignments obtained by BLAST. All sequences were manually inspected and corrected for accuracy and completeness. We excluded from further analysis a few DNA sequences containing ambiguous nucleotides. Whenever there were distinct annotations in GenBank for the start codon of the same *inc* gene in *C. trachomatis* archetype ocular (A/Har13), urogenital (D/UW3), and LGV (L2/434) strains, we used: for *ct036* and *ct119*/*incA*, in all cases the start codons of each of the genes as annotated for D/UW3 and A/Har13; for *ct115*/*incD* and *ct192*, the start codons of each of the genes as annotated for A/Har13, D/UW3, or L2/434, in ocular, urogenital or LGV strains, respectively; for *ct226*, in all cases the start codon as annotated for A/Har13 (see Supplemental Table S3.1 and S3.3).

3.3.5. Sequence alignments, and analyses of polymorphisms, phylogeny, and molecular evolution

Alignments of the amino acid sequences of the Inc, Pmp, and housekeeping proteins, deduced from the retrieved nucleotide sequences, were generated using the ClustalW algorithm in MEGA5 software (www.megasoftware.net) (233). All alignments were manually inspected and corrected for artifacts. We excluded all strain-specific pseudogenes (Supplemental Table S3.3) from further phylogenomic and evolutionary analyses.

For the analyses of polymorphism, phylogeny, and molecular evolution, various tools present in MEGA5 were used, essentially as previously described (111). Briefly, for analyses of polymorphism, we computed pairwise, overall, within groups (ocular, urogenital, or LGV), and between groups (ocular *vs* urogenital, ocular *vs* LGV, or urogenital *vs* LGV) amino acid *p*-distances. For analyses of phylogeny, trees were generated using the Neighbor-Joining method (234). The generated phylograms of Inc proteins were inspected for separate branches (segregation) of all ocular, urogenital, or LGV strains. This analysis was supported by comparison of pairwise amino acid *p*-distances between and within disease groups. The phylogeny of an Inc protein was considered to segregate a disease group if the maximum pairwise amino acid *p*-distance within groups was less than any of the pairwise amino acid *p*-distances between groups. For molecular evolution analyses, we used the Kumar method (39) to compute overall means of non-synonymous (d_N) and synonymous (d_S) substitutions (per non-synonymous or synonymous sites, respectively) and to find genes that may be under positive selection. By using the codon based Z-test of selection in MEGA5, genes were considered under positive selection if they showed a statistically significant value ($P < 0.05$) to reject the null hypothesis of strict-neutrality ($d_N = d_S$) in favour of both positive selection ($d_N > d_S$) and lack of neutrality ($d_N \neq d_S$). All these analyses were performed by Bootstrap with 1,000 replicates, and with the pairwise-deletion option selected.

3.3.6. Real-time quantitative PCR

The expression of *inc* genes during the developmental cycle of *C. trachomatis* B/Har36, C/TW3, E/Bour, F/CS465-95, L2/434, L2b/CS19-08, and L3/404 was estimated by determining *inc* mRNA levels at different times post-infection by real-time quantitative PCR (RT-qPCR). These experiments were essentially done as previously described (231,277). Briefly, for each strain, 6 tissue culture flasks of 25 cm^2 containing monolayers of HeLa 229 cells were inoculated at a multiplicity of infection of 1; cells were harvested at 2, 6, 12, 20, 30 and 42 h post-infection, by scrapping in ice-cold phosphate buffered saline. The cell suspension was sonicated to disrupt mammalian cells and promote bacterial release, followed by low-speed centrifugation at 4°C. The supernatant was then frozen in liquid nitrogen and stored at -80ºC. These samples were used for total RNA purification and generation of cDNA, as previously described (231). Primers (available upon request) were designed for each *inc* gene using Primer Express (Applied Biosystems), based on identical *C. trachomatis* sequences between strains. The RT-qPCR assays were done using the ABI 7000 SDS, SYBR Green chemistry and optical plates (Applied Biosystems), as previously described (231,277). At each time-point, raw RT-qPCR data of each *inc* gene was normalized against the data obtained for the *16SrRNA* transcript, as it was previously demonstrated that this is a good endogenous control (231). The final results were based on at least two independent experiments.

3.3.7. Transcription linkage analysis and identification of transcriptional start sites

We searched for disease group-specific nucleotides within the promoter region of 3 *inc* genes (*ct058*, *ct192*, and *ct214*) that showed differential gene expression. For this, we determined their transcriptional start sites (TSSs). For *ct058* and *ct192*, it was ambiguous whether their promoter would lie immediately upstream from their predicted start codons (Figure 3.5A). Therefore, we used reverse transcription coupled with PCR (RT-PCR) to determine if *ct058* and *ct059*, or *ct192* and *ct193*, are part of the same transcriptional unit. For this, RNA was isolated from HeLa 229 cells infected for 30 h with *C. trachomatis* L2/434 using NZY Total RNA kit (NZYTech). cDNA was then generated by using random hexamers and iSCRIPT (BioRad). Primers (see Supplemental Table S3.2) were designed to generate PCR products containing ~300 bp upstream and downstream from the predicted start codons. PCR products were obtained using DreamTaq DNA polymerase (MBI Fermentas). As controls for the PCRs, we also used as template: the product of a typical reverse transcription reaction but without iSCRIPT; total DNA, isolated using the NZY Tissue gDNA kit (NZYTech), either from cells infected with strain L2/434 for 42 h or left uninfected.

The identification of the TSSs of *ct059* (upstream from *ct058*, and in the same transcriptional unit; Figure 3.5A), *ct192*, and *ct214* in L2/434 was done by 5' rapid amplification of cDNA ends (RACE), using the 5'/3'RACE kit, 2nd Generation (Roche). We used RNA isolated as described above from HeLa 229 cells infected for 30 h with *C. trachomatis* L2/434 and the primers listed in Supplemental Table S3.2. Final PCR amplification of double stranded cDNA was done with Phusion DNA polymerase (Finnzymes). PCR products were purified after agarose gel electrophoresis, using High Pure PCR Purification Kit (Roche), and then subjected to DNA sequencing. To analyze the determined TSSs in the context of the promoter regions of *ct059-ct058*, *ct192*, and *ct214* in all strains

used in the RT-qPCR assays, the corresponding nucleotide sequences were either retrieved from GenBank (E/Bour, L2/434, and L3/404) (Supplemental Table S3.1) or determined by DNA sequencing (C/TW3, B/Har36, F/CS465-95, and L2b/CS19-08), as previously described (111) and using the primers listed in the Supplemental Table S3.2. All these manipulations were done according to instructions from the indicated manufacturers.

3.3.8. Nucleotide sequence accession numbers

The sequences of the promoter regions of *ct192*, and *ct214* and of *ct059-ct058* in C/TW3, B/Har36, F/CS465-95, and L2b/CS19-08 determined in this study were submitted to GenBank and are available under accession numbers JX451863 - JX451874.

3.4. Results

3.4.1. Identification of T3S signals in *C. trachomatis* Inc proteins

We focused on 48 predicted *C. trachomatis* Inc proteins (i.e. possessing a bilobed hydrophobic domain), which have been singled out and studied by Li et al (300) (Table 3.1). Thus far, 23 of these proteins have been detected in the inclusion membrane by immunofluorescence microscopy using specific antibodies (known Inc proteins) [references (129,258,300), and references therein], but not the other predicted Inc proteins (putative Inc proteins) (129,292) (Table 3.1).

Table 3.1. Summary of T3S signals found in known and putative Inc proteins of *C. trachomatis* analyzed in this work.

Inc protein[a]	T3S signal	Reference
Known [b]		
CT101	No	This work
CT115/IncD	Yes	(132); this work
CT116/IncE	Yes	(132)
CT117/IncF	Yes	This work
CT118/IncG	Yes	(132)
CT119/IncA	Yes	(132); this work
CT147	ND [e]	-
CT222	Yes	This work
CT223	Yes	(132)
CT225	No	This work
CT226	Yes	(129)
CT228	Yes	(129); this work
CT229	Yes	(132)
CT232/IncB	Yes	This work
CT233/IncC	Yes	(130,132); this work
CT249	Yes	(129); this work
CT288	Yes	(132)
CT358	Yes	(129); this work
CT440	Yes	(129); this work
CT442	Yes	(132)
CT618	Yes	This work

CT813	Yes	This work
CT850	Yes/No[d]	(129); this work
Putative [c]		
CT005	Yes	This work
CT006	ND	-
CT036	Yes	This work
CT058	Yes	(129); this work
CT134	No	This work
CT135	Yes	This work
CT164	No	This work
CT179	No	This work
CT192	No/Yes[d]	(129); this work
CT195	Yes	(129); this work
CT196	Yes	This work
CT214	Yes	This work
CT224	Yes	This work
CT227	Yes	This work
CT300	Yes	This work
CT345	Yes	This work
CT357	Yes	This work
CT365	Yes	This work
CT383	Yes	(129); this work
CT449	Yes	This work
CT483	Yes	This work
CT484	No	(129); this work
CT565	No	(129); this work
CT728	No	This work
CT789	Yes	This work

[a] Proteins containing a bilobed hydrophobic motif that were analyzed by Li et al (300), which we selected to study in this work. We did not consider proteins (Cap1 and CopN) which localize to the inclusion membrane but which do not possess the bilobed hydrophobic domain (160,291). More recent bioinformatics-based analyses identified additional putative Inc proteins (CT018, CT079, CT081, CT244, CT324, CT326, CT556, CT578, CT616, CT618, CT642, CT645, CT788, CT789, CT814.1, CT819, CT837, CT846, and CT873) in *C. trachomatis* (129,292), but these proteins were not analyzed in our study.

[b] Known Inc proteins contain a bilobed hydrophobic motif and have been localized to the inclusion membrane by immunofluorescence microscopy using specific antibodies [references (129,258,300), and references therein].

[c] Putative Inc proteins contain a bilobed hydrophobic motif but have not yet been localized to the inclusion membrane.

[d] Conflicting data between our observations and previous analyses using *S. flexneri* as a heterologous bacterial host.

[e] ND, not determined.

As Inc proteins are believed to be transported into the inclusion membrane by a T3S mechanism (129,130,132,289), we used *Y. enterocolitica* as heterologous bacteria to identify a possible N-terminal T3S signal in putative Inc proteins of *C. trachomatis* by comparison to known Inc proteins. We sought to obtain additional indirect evidence (besides the bilobed hydrophobic motif) that putative Inc proteins localize to the inclusion membrane. We analyzed secretion of hybrid proteins comprising the first 20 amino acids of each putative or known Inc and TEM-1 by T3S-proficient (ΔHOPEMT) or T3S-deficient (ΔHOPEMT ΔYscU) *Y. enterocolitica* (Figure 3.1). The different *Y. enterocolitica* strains were incubated in T3S-inducing conditions (296), followed by fractionation of the bacterial cultures into culture supernatants and bacterial pellets and subsequent immunoblotting analyses of the proteins in the two fractions (examples in Figure 3.1A and 3.1B).

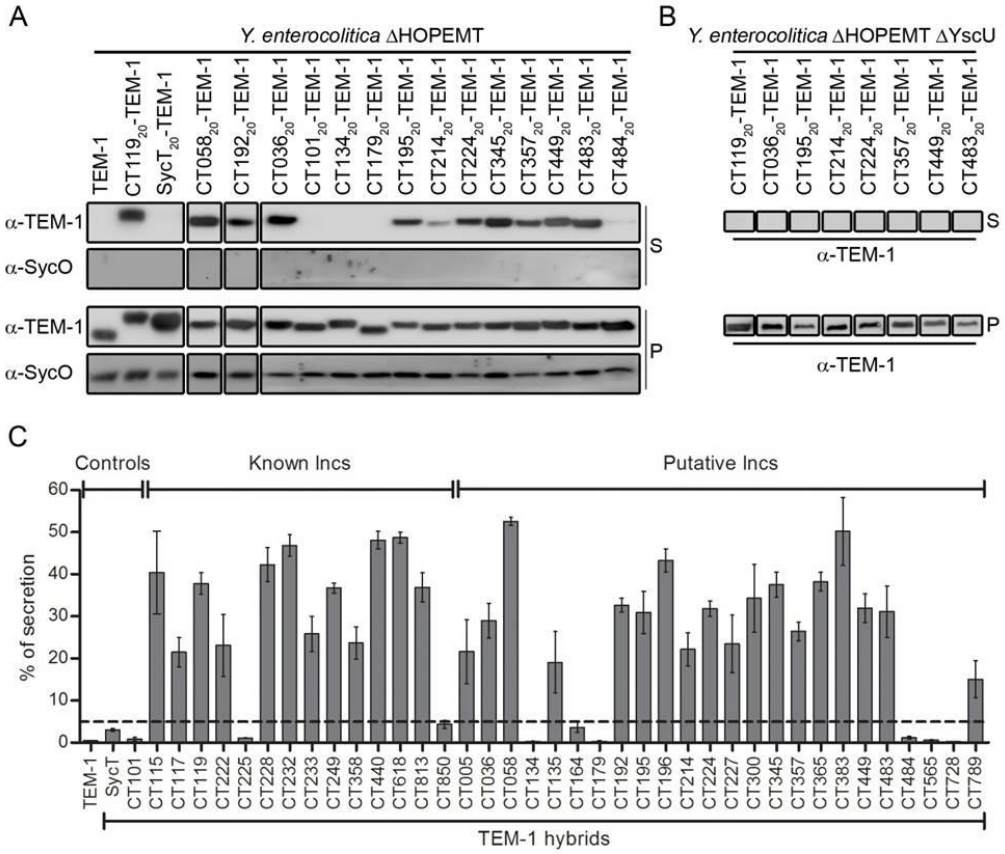


Figure 3.1. Type III secretion (T3S) signals in *C. trachomatis* Inc proteins. *Y. enterocolitica* T3S-proficient (ΔHOPEMT) and T3S-defective (ΔHOPEMT ΔYscU) bacteria were used to analyse secretion of hybrid proteins comprising the first 20 amino acids of Inc proteins, or of *Y. enterocolitica* SycT, fused to the mature form of TEM-1 β-lactamase (TEM-1). (A) and (B) Immunoblots show the result of representative assays in which proteins in culture supernatants (S - secreted proteins), and in bacterial pellets (P - non-secreted proteins) from ~5x10^7 bacteria were loaded per lane. SycT and SycO are strictly cytosolic *Yersinia* T3S chaperones (295,299). SycT$_{20}$-TEM-1 was a negative control for the T3S assays. Immunodetection of SycO ensured that presence of TEM-1 hybrid proteins in culture supernatants was not a result of bacterial lysis or contamination. (C) The percentage (%) of secretion of each TEM-1 hybrid was calculated by densitometry, as the ratio between the amount of secreted and total protein. The threshold to decide whether a protein was secreted was set to 5% (dashed line), based on the % of secretion of SycT$_{20}$-TEM-1. Data are the mean ± SEM from at least 3 independent experiments. Please note that in this work we did not analyse all described putative Inc proteins (see legend of Table 3.1) and that T3S signals were not analysed for all known Inc proteins.

In total, we have analyzed T3S signals in 24 putative Inc proteins (18 have not been previously analyzed for T3S) and in 15 known Inc proteins (7 have not been previously analyzed for T3S) (Figure 3.1C and Table 3.1). The expression levels of the TEM-1 hybrid of putative Inc CT006 were extremely low, which hampered the analysis of a T3S signal in this protein (data not shown). These experiments led to the identification of a T3S signal in 18 putative Inc proteins and in 12 known Inc proteins (Figure 3.1C and Table 3.1). This revealed 5 known and 14 putative Inc proteins as novel *C. trachomatis* T3S substrates (Figure 3.1C and Table 3.1). However, we did not detect a clear T3S

signal in three known Inc proteins (CT101, CT225, and CT850; Figure 3.1 and Table 3.1). Their T3S signal could extend beyond the first 20 amino acids or might not be recognized by the *Y. enterocolitica* T3S system; alternatively, they may be transported into the inclusion membrane by a distinct mechanism. Regardless of the exact explanation, this implies that the lack of a detectable T3S signal could not be taken as a definitive indication that a putative Inc does not localize to the inclusion membrane. Furthermore, the percentage of putative and known Inc proteins analyzed that displayed an N-terminal region recognized by the *Y. enterocolitica* T3S machinery was nearly identical [75% (18 out of 24) and 80% (12 out of 15), respectively]. Overall, these analyses indicated that most of the putative Inc proteins analyzed are T3S substrates.

3.4.2. Differences in the amino acid sequences of Inc proteins among *C. trachomatis* strains correlate with the type of infection and with tissue tropism

The nucleotide sequences of the genes encoding the selected 48 known and putative Inc proteins were retrieved from 51 fully sequenced *C. trachomatis* genomes (7 ocular, 23 urogenital, and 21 LGV strains) (Supplemental Table S3.1). In an initial analysis, 7 *inc* genes proved to be pseudogenes in different *C. trachomatis* strains (*ct058*, *ct101*, *ct135*, *ct192*, *ct227*, *ct228*, *ct300* (Supplemental Table S3.1). *ct358* was described as a pseudogene in L2/434 and L2/UCH-1 (46), and its nucleotide sequence is 100% identical in all LGV strains. However, analysis of its nucleotide sequence suggests that *ct358* may encode a functional protein that is 7 amino acids shorter at its C-terminus than CT358 (178 amino acid residues long) in ocular and urogenital strains (Supplemental Table S3.1). In all the seven *inc* genes that we identified as pseudogenes the full-length gene is disrupted by a mutation that leads to a significantly truncated protein. The only disease group-specific correlation was observed with *ct300*, which is a pseudogene in all LGV strains analyzed. Therefore, the encoded protein is expendable for LGV infections. In addition, *ct058* and *ct101* revealed to be pseudogenes in almost all analyzed ocular and urogenital strains, respectively (Supplemental Table S3.1). *ct135*, *ct192*, *ct227*, and *ct228*, are only pseudogenes in a few strains, with no obvious correlation with ocular, urogenital, or LGV disease groups (see Supplemental Table S3.1). Furthermore, 12 *inc* genes showed small deletion and insertion events (Supplemental Table S3.1).

To understand if the amino acid sequence of Inc proteins vary among strains, we determined the overall mean genetic distance (amino acid *p*-distance) for each Inc among all 51 *C. trachomatis* strains (discarding strain-specific pseudogenes) (Figure 3.2A). As reference, we also analyzed the 9 Pmps and 9 housekeeping proteins of *C. trachomatis* previously shown to be polymorphic (163) (Figure 3.2A). The Pmps should localize to the bacterial outer membrane and the housekeeping proteins within the bacterial cell. The average *p*-distance was 0.017 (SEM, 0.002) for Incs, 0.020 (SEM, 0.007) for Pmps, and 0.013 (SEM, 0.003) for housekeeping proteins. Based on this, and considering the average *p*-distance for Incs as a cut-off value, we defined 19 (40% of the total) Inc proteins as “polymorphic” (*p*-distance ≥ 0.017) (Figure 3.2A). As comparison, 4 Pmps (44%) and 2 housekeeping proteins (22%) displayed a *p*-distance of ≥ 0.017. This showed that the overall degree

of polymorphism in Inc proteins among *C. trachomatis* strains is similar to that of Pmps and is higher than that of known polymorphic housekeeping proteins.

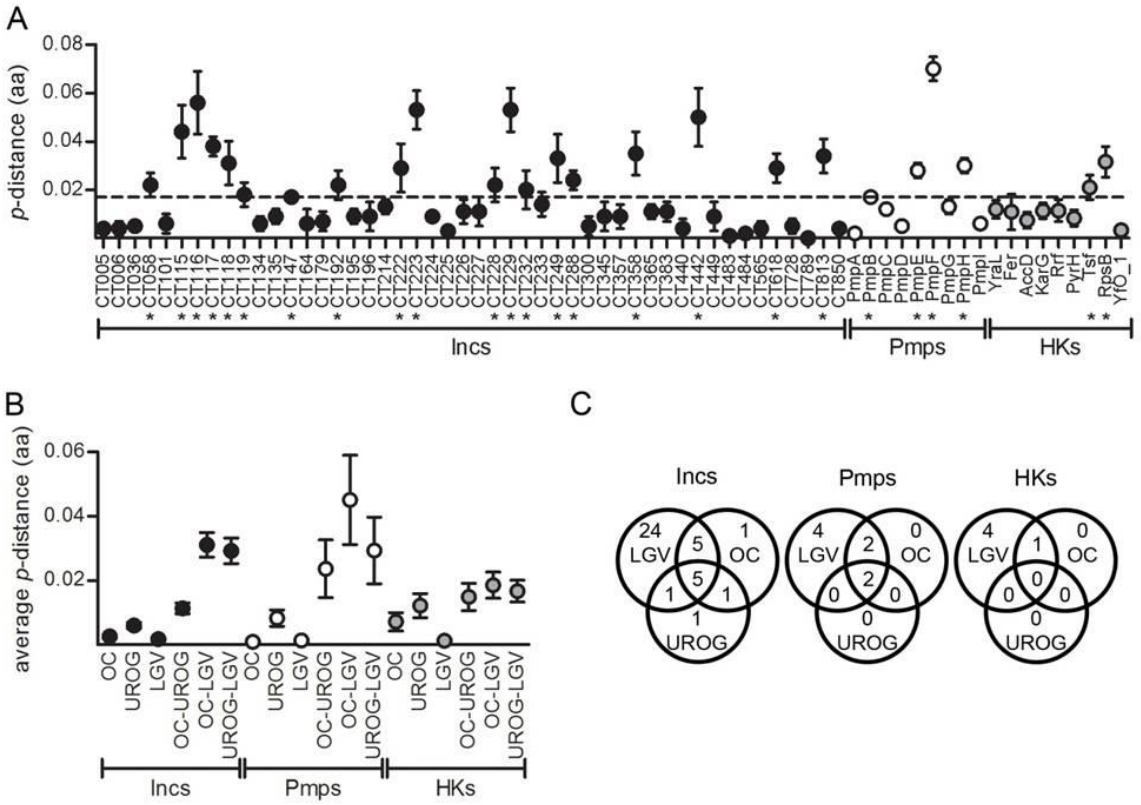


Figure 3.2. Polymorphisms in *C. trachomatis* Inc proteins. Polymorphic membrane proteins (Pmps) and housekeeping proteins (HKs) were analyzed as reference. (A) Overall mean genetic distance (polymorphisms) based on the *p*-distance between all possible pairs of amino acid (aa) sequences of Inc proteins, Pmps, and HKs among *C. trachomatis* strains. Proteins marked with an asterisk (*) have an aa *p*-distance that is equal or higher than the average value for Inc proteins (0.017; shown as a dashed line). (B) Average mean genetic distance, based on the *p*-distance between all possible pairs of aa sequences within (ocular, urogenital, or LGV; OC, UROG, or LGV, respectively) or between (OC-UROG, OC-LGV, or UROG-LGV) *C. trachomatis* disease groups. (C) Venn diagrams showing the phylogenetic segregation of *C. trachomatis* disease groups based on Neighbor-Joining trees of Inc proteins, Pmps or HKs and on pairwise aa *p*-distances between and within disease groups, for all possible pairs of Inc proteins, Pmps or HKs sequences from the *C. trachomatis* strains analysed. All these analyses were performed by Bootstrap with 1,000 replicates. Error bars represent SEM.

To understand if the amino acid differences between Inc proteins were related with the type of infection and with tissue tropism, we determined the average *p*-distance of Inc proteins within and between the three groups of strains (ocular, urogenital, and LGV) (Figure 3.2B). This showed that the differences were largely due to variations between Inc proteins from LGV strains and ocular (average *p*-distance = 0.031; SEM, 0.004) or urogenital strains (average *p*-distance = 0.029; SEM, 0.004) (Figure 3.2B). These average *p*-distances were significantly higher (in all cases, $p < 0.0001$; two-tailed

t-test) than those between Inc proteins from ocular and urogenital strains (average *p*-distance = 0.011; SEM, 0.002) or within Inc proteins from the same disease groups (average *p*-distance < 0.006) (Figure 3. 2B). Similar observations were made for Pmps except that, in contrast to Inc proteins, the average *p*-distance between Pmps from ocular and urogenital strains was not significantly different than that between Pmps from LGV strains and ocular or urogenital strains (in both cases, p > 0.05; two-tailed *t*-test) (Figure 3.2B). Also in contrast to the Inc proteins, the variation in the amino acid sequences of housekeeping proteins between each of the three groups were clearly not different (in all cases, p > 0.05; two-tailed *t*-test) (Figure 3.2B). Furthermore, the amino acid sequences of housekeeping proteins varied nearly as much between or within groups (Figure 3.2B). The main exception was the LGV group, within which the housekeeping proteins, as Inc proteins and Pmps, showed to be extremely conserved (average *p*-distance < 0.002 for the three cases) (Figure 3.2B). Thus, the separation (*p*-distance) between LGV strains and ocular or urogenital strains is much more marked for Inc proteins than for Pmps or housekeeping proteins.

We then made and analyzed phylogenetic reconstructions based in the amino acid sequence of Inc proteins. The phylograms of five Inc proteins (10%) showed tropism, i.e. segregation of the three disease groups, and those of 38 Inc proteins (84%) evidenced segregation of at least one disease group (Figure 3.2C and Supplemental Table S3.4). The phylograms of a total of 35 Inc proteins (73%) showed segregation of LGV strains, while only 12 (25%) and 8 (17%) displayed clustering of ocular and urogenital strains, respectively (Figure 3.2C and Supplemental Table S3.4). This scenario was mirrored by the phylograms of Pmps but was in contrast with the phylograms of housekeeping proteins, in which segregation by disease group was less often seen (Figure 3.2C; see also Supplemental Table S3.4).

In summary, there are significant differences in the amino acid sequences of Inc proteins among *C. trachomatis* strains and they correlate with the type of infection and with tissue tropism. In particular, differences are mostly between Inc proteins from LGV strains and Inc proteins from ocular or urogenital strains.

3.4.3. *inc* genes of *C. trachomatis* have distinct evolutionary dynamics and several *inc* genes are likely under positive selection

To understand the underlying evolutionary pressures that drive amino acid changes in Inc proteins, we analyzed the molecular evolution of *inc* genes. We first determined overall d_N/d_S values for *inc* genes, by comparison to the 9 *pmp* genes and the 9 selected housekeeping genes of *C. trachomatis*. We found that 24 *inc* genes (50%) showed a d_N/d_S value which was > 1, and that in 4 *inc* genes all substitutions were non-synonymous (Figure 3.3A and Supplemental Table S3.5). In contrast, only 2 *pmp* genes (22%) showed a d_N/d_S value that was > 1, and all housekeeping genes displayed a d_N/d_S value of < 1 (Figure 3.3A and Supplemental Table S3.5).

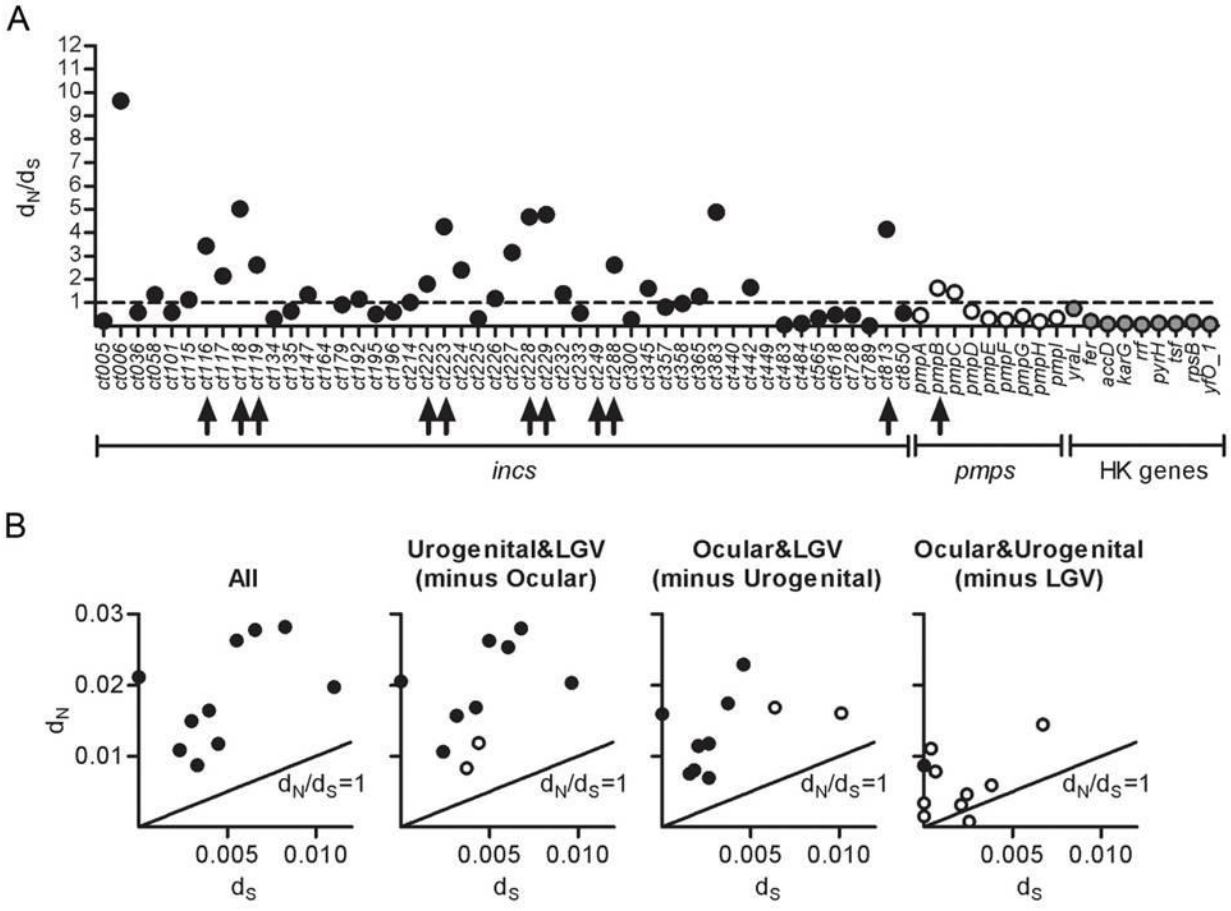


Figure 3.3. Evolutionary dynamics of *inc* genes. Genes encoding polymorphic membrane proteins (*pmps*) or housekeeping (HK) proteins were used as reference. (A) Ratio of non-synonymous (d_N) to synonymous (d_S) substitutions among *C. trachomatis* strains. The dashed line indicates neutrality (d_N/d_S = 1). The arrows specify genes likely under positive selection, according to the codon based Z-test of selection (see Materials and Methods). (B) Distribution of d_N *vs* d_S values for the 10 *inc* genes likely under positive selection, comparing the impact of artificially discarding ocular, urogenital, or LGV strains relative to the analysis with all *C. trachomatis* strains. The straight line in each graph indicates neutrality (d_N/d_S = 1). In all cases, *inc* genes likely under positive selection (codon based Z-test of selection) are depicted as black circles, while *inc* genes for which statistical support of likely positive selection can no longer be detected after discarding a particular group of strains are depicted as white circles. All these analyses were performed by Bootstrap with 1,000 replicates. For sake of clarity, SEM for values in both panels are presented only in Supplemental Table S3.5.

Analyses of the overall d_N/d_S values using the codon based Z-test of selection (see Materials and Methods) yielded statistically significant values for 10 *inc* genes (*ct116*/*incE*, *ct118*/*incG*, *ct119*/*incA*, *ct222*, *ct223*, *ct228*, *ct229*, *ct249*, *ct288*, and *ct813*), but only for one *pmp* gene and for no housekeeping gene (Figure 3.3A and Supplemental Table S3.5). This indicated that these 10 *inc* genes are likely under positive selection. We then aimed to understand which group of strains might cause the detection of this possible evolutionary trend. For this, we assessed the impact of artificially

removing ocular, urogenital, or LGV strains from the analyses (Figure 3.3B). This showed that discarding LGV strains caused major alterations on the d_N and d_S values and confined the ability to detect Z-test-based likely positive selection to only one *inc* gene, whereas discarding ocular or urogenital strains had less pronounced effects (Figure 3.3B). In addition, within each of the 10 proteins encoded by *inc* genes likely under positive selection, we found 113 amino acid residues that are disease group-specific (Table 3.2). Among these residues, 104 (92%) were in Inc proteins from LGV strains, 77 (74%) of which localized in regions of the proteins predicted to be on the cytoplasmic side of the inclusion membrane (Table 3.2).

Overall, this suggested that the evolutionary dynamics of *inc* genes is distinct from *pmps* or housekeeping genes, and that LGV-specific amino acid residues in a subset of Inc proteins might be involved in the unique ability of *C. trachomatis* LGV strains to infect macrophages and disseminate into lymph nodes.

Table 3.2. Amino acids residues within Inc proteins encoded by genes likely under positive selection that are specific of *C. trachomatis* disease groups.

Inc	Length (aa)[a]	TM segments[b]	LGV[c]	Urogenital[c]	Ocular[c]
CT116/IncE	132	[36-59], [64-87]	V37**A**, V38**A**, S39**C**, S64**G**, I68**L**, V72**I**, I84**T**, D92**N**	D19**G**	
CT118/IncG	167	[33-57], [63-88]	A33**V**, F77**C**, C79**Y**, N87**S**, F135**L**, G136**R**, H165**R**		
CT119/IncA	273	[35-59], [64-84]	I75**T**, Q207**E**, V211**A**, V212**A**		
CT222	129	[39-63], [69-93]	Y10**C**, I75**-**, C89**Y**, L125**I**, V126**S**, Y127**V**, S128**F**, N129**H**		
CT223	270	[38-61], [67-91]	G11**R**, A65**T**, K100**R**, I108**L**, K127**G**, N130**D**, P134**-**, C153**Y**, E160**D**, T166**K**, H200**Y**, E204**D**, N207**R**, L208**M**, R231**L**, V241**A**, P251**L**, D261**Y**, G263**-**	L225**F**	A40**V**, A75**V**, D142**G**, V161**M**
CT228	196	[38-59], [65-86]	I22**T**, A68**V**, A72**P**, C90**Y**, V119**A**, A143**V**, I144**M**, V146**F**		
CT229	215	[42-65], [71-90]	A39**S**, V99**I**, G112**E**, K114**E**, F121**S**, Q125**R**, V126**A**, H137**Y**, Q144**K**, Y152**H**, E154**A**, E164**K**, G173**R**, S181**N**, T201**A**	Q102**H**, I156**V**	C157**S**
CT249	116	[51-72], [78-97]	H8**Y**, D24**N**, V80**T**, A89**I**		
CT288	563	[36-58], [65-88], [242-263], [269-291]	A67**T**, I69**V**, E102**A**, S121**P**, A194**T**, S198**T**, N203**K**, T209**I**, A258**T**, I260**V**, I261**V**, L318**W**, S346**G**, L388**V**, -474**D**, A516**P**, V531**L**, A532**T**		
CT813	264	[41-61], [68-94]	N7**T**, V76**I**, E107**K**, K125**E**, R128**Q**, T145**A**, E161**K**, A163**G**, E168**K**, E171**K**, E172**Q**, I181**V**, I236**T**		

[a] Based on the protein sequence annotation of the *C. trachomatis* strain D/UW3.
[b] Position of transmembrane (TM) domains in the corresponding Inc proteins, obtained from (129), except for CT223, for which we found only the two indicated TM domains.
[c] Residues specific to Inc proteins from each disease group are indicated in bold, relative to the amino acid in the same position in the other disease groups, and those in regions predicted to be on the cytoplasmic side of the inclusion membrane are underlined (we considered that the loop region within two TM segments faces the lumen of the inclusion); a dash indicates a deletion or an insertion.

3.4.4. Disease group-specific expression of *C. trachomatis inc* genes

To analyze if there were differences in the expression of *inc* genes between *C. trachomatis* strains that correlate with the type of infection or with tissue tropism, we used RT-qPCR to determine the mRNA levels of the 48 selected *inc* genes throughout the developmental cycle of *C. trachomatis*. We aimed to find *inc* genes showing differences in the highest mRNA levels during the cycle (peak of expression) or in the variation of mRNA levels throughout development (profile of expression) between *C. trachomatis* strains.

We first infected HeLa 229 cells with ocular (C/TW3), urogenital (E/Bour), or LGV (L2/434) prototype strains. Total RNA was isolated at 2, 6, 12, 20, 30, and 42 h post-infection, which was used to generate cDNA for RT-qPCR assays (complete data are shown in Supplemental Table S3.6). Generally, the comparison of the peak of expression revealed differences from 30 to 60-fold (depending on the strain) between *inc* genes (Figure 3.4A). However, the average of the peaks of expression of *inc* genes in C/TW3, E/Bour, or L2/434 were not significantly different between each strain (Figure 3.4A). Regarding the profile of expression, essentially as previously described (5, 41, 52), we have identified: *inc* genes with highest expression at 2 or 6 h post-infection that then either decreased or remained constant throughout the cycle (early-cycle genes); *inc* genes with highest expression only at 12 or 20 h post-infection and which then decreased or remained constant at later time points (mid-cycle genes); and *inc* genes with highest expression only at 30 or 42 h post-infection (late-cycle genes) (Figure 3.4B and examples in Supplemental Figure S3.1). We also identified 4 *inc* genes in E/Bour and 5 *inc* genes in L2/434 that were simultaneously “early- and late-cycle genes”, showing identically high mRNA levels both at 2 or 6 h post-infection and at 30 or 42 h post-infection but lower expression at mid-cycle (Figure 3.4B). This was typically the case of *ct214* and *ct288* in L2/434 (see Supplemental Figure S3.2). Generally, 35 *inc* genes showed the same profile of expression in the three *C. trachomatis* strains, and the majority of *inc* genes showed an early-cycle profile of expression (34 in C/TW3, 27 in E/Bour, and 32 in L2/434) (Fig. 4B). In spite of these common features, we identified 9 *inc* genes (*ct005*, *ct058*, *ct192*, *ct214*, *ct232*/*incB*, *ct249*, *ct288*, *ct440*, and *ct442*) whose peak of expression consistently showed differences of > 2-fold between strains and/or whose profile of expression displayed differences that could not be explained by distinct growth kinetics of the strains (Supplemental Figure S3.2). With the exception of *ct214*, all these genes showed consistently higher peaks of expression in L2/434 than in C/TW3 and/or E/Bour; *ct442* also showed a higher peak of expression in E/Bour than in C/TW3 (Supplemental Figure S3.2).

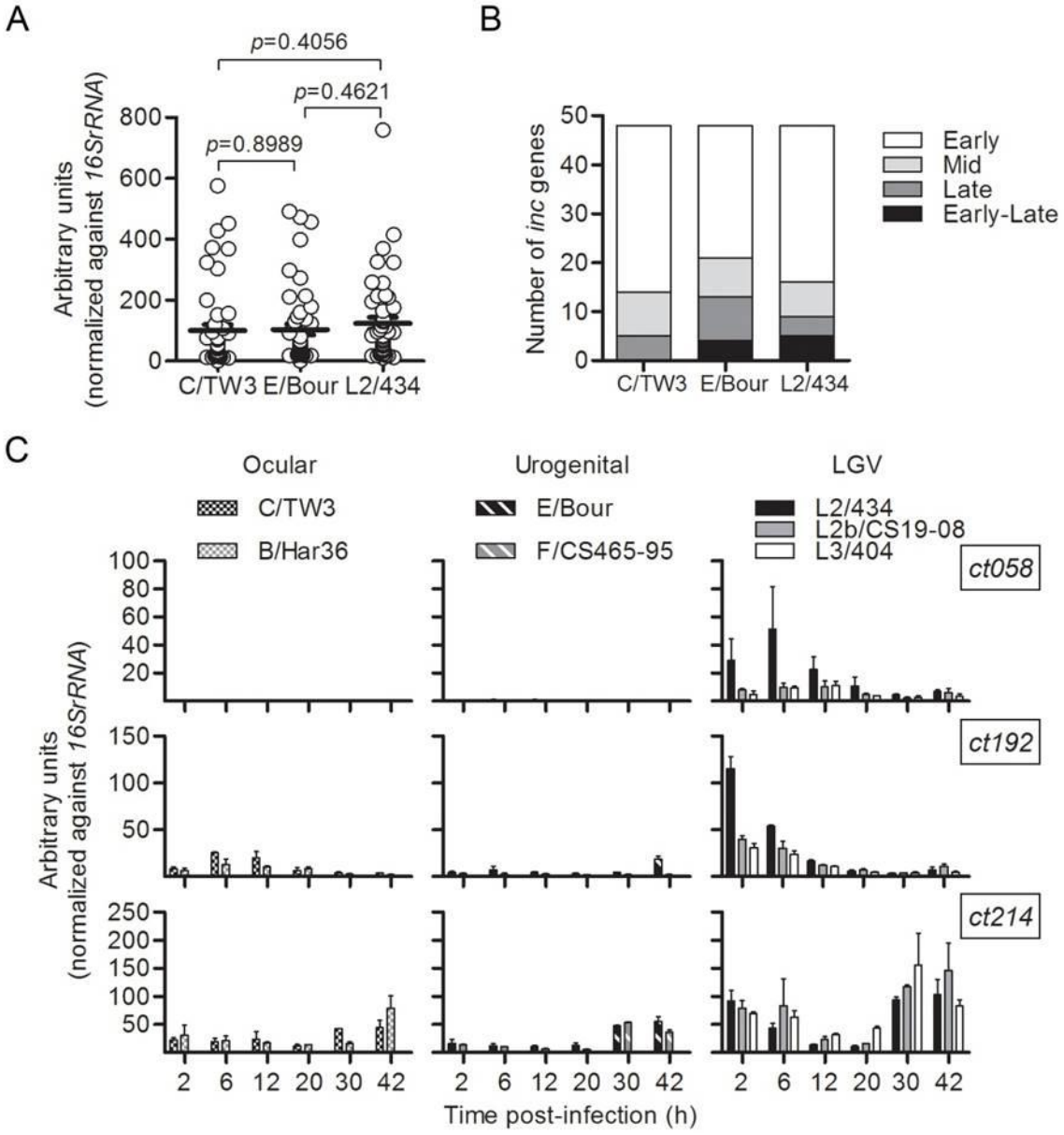


Figure 3.4. mRNA levels of *inc* genes during the developmental cycle of different *C. trachomatis* strains. The mRNA levels of 48 *inc* genes (A and B), or of *ct058*, *ct192*, and *ct214* (C), were analyzed by RT-qPCR throughout the developmental cycle of the indicated prototype (B/Har36, C/TW3, E/Bour, L2/434, and L3/404) and clinical (F/CS465-95, and L2b/CS19-08) strains. (A) Peak of expression (highest mRNA levels during the developmental cycle) of each *inc* gene. The *p* values were calculated by two-tailed *t*-tests. (B) Number of *inc* genes showing the indicated profiles of expression (variation of mRNA levels during the developmental cycle). (C) The expression values (mean ± SEM) result from the RT-qPCR raw data ($x10^5$) of each gene normalized against the *16SrRNA*, from at least two independent experiments. Complete data are shown in Supplemental Table S3.6.

To analyze whether the differences found in the expression of *ct005*, *ct058*, *ct192*, *ct214*, *ct232*/*incB*, *ct249*, *ct288*, *ct440*, and *ct442* were strain-specific or disease group-specific, we determined the mRNA levels of these genes during the developmental cycle of additional *C.*

trachomatis strains. For this, we infected HeLa 229 cells with ocular B/Har36 (prototype), urogenital F/CS465-95 (clinical isolate), LGV L2b/CS19-08 (clinical isolate), or LGV L3/404 (prototype) strains. The infected cells were processed for RT-qPCR assays and analyzed as described above (complete data are shown in Supplemental Table S3.6). We detected disease group-specific differences in gene expression only for *ct058*, *ct192*, and *ct214* (Figure 3.4C): *ct058* showed an "early-cycle gene" profile of expression in which mRNA levels were evident for LGV strains but only vestigial for ocular and urogenital strains; *ct192* showed only a clear "early-cycle gene" profile of expression in LGV strains and its expression levels were generally higher in LGV strains than in ocular or urogenital strains; *ct214* displayed an "early- and late-cycle gene" profile of expression in LGV strains but a late-cycle gene profile of expression in ocular or urogenital strains. Therefore, we have identified 3 *inc* genes (*ct058*, *ct192*, and *ct214*) showing differences in gene expression between *C. trachomatis* strains that correlate with the type of infection and tissue tropism, in particular with LGV isolates.

3.4.5. Identification of LGV-specific nucleotides in the promoter regions of *ct058*, *ct192*, and *ct214*

We next attempted to obtain insights on the genetic basis for the disease group-specific expression of *ct058*, *ct192* and *ct214* by analyzing the promoter region of these genes in *C. trachomatis* L2/434. The gene organization of these *inc* genes suggested that the promoter region of *ct214* should lie between the start codons of *ct214* and *ct215* or within the first codons of *ct215* (Figure 3.5A). However, the localization of the promoter regions of *ct058* or *ct192* was unclear, as these genes could be co-transcribed with *ct059* or *ct193*, respectively (Figure 3.5A). Therefore, we used RT-PCR with cDNA template generated from total RNA of HeLa 229 cells infected with *C. trachomatis* L2/434 to determine the possible transcriptional linkage between *ct058* and *ct059*, and between *ct192* and *ct193*. This indicated that *ct058* is co-transcribed with *ct059*, which likely encodes a ferredoxin, and transcription of *ct192* is unlinked from *ct193* (Figure 3.5A and 3.5B). We previously detected clear measurable mRNA levels of *ct059* in *C. trachomatis* ocular and urogenital strains (231). We re-confirmed this, using the same biological samples in which the levels of *ct058* mRNA were vestigial (data not shown). Therefore, we tentatively propose that expression of *ct058* in ocular and urogenital strains might be down-regulated by specific 3' to 5' post-transcriptional processing of the *ct059-ct058* transcript.

To precisely define the promoter region of *ct059-ct058*, *ct192*, and *ct214*, we determined their TSSs by RACE, using as template total RNA of cells infected with *C. trachomatis* L2/434 and primers complementary to the *ct059*, *ct192*, or *ct214* mRNA (Figure 3.5A and 3.5C-E). When using a primer complementary to the *ct058* mRNA, we were unable to identify a *ct058*-exclusive TSS upstream from its start codon in L2/434 (data not shown). The TSS of *ct214* matched the one previously identified by deep-sequencing in strain L2b/UCH-1 (102), while the TSSs of *ct059-ct058* and *ct192* have not been identified before.

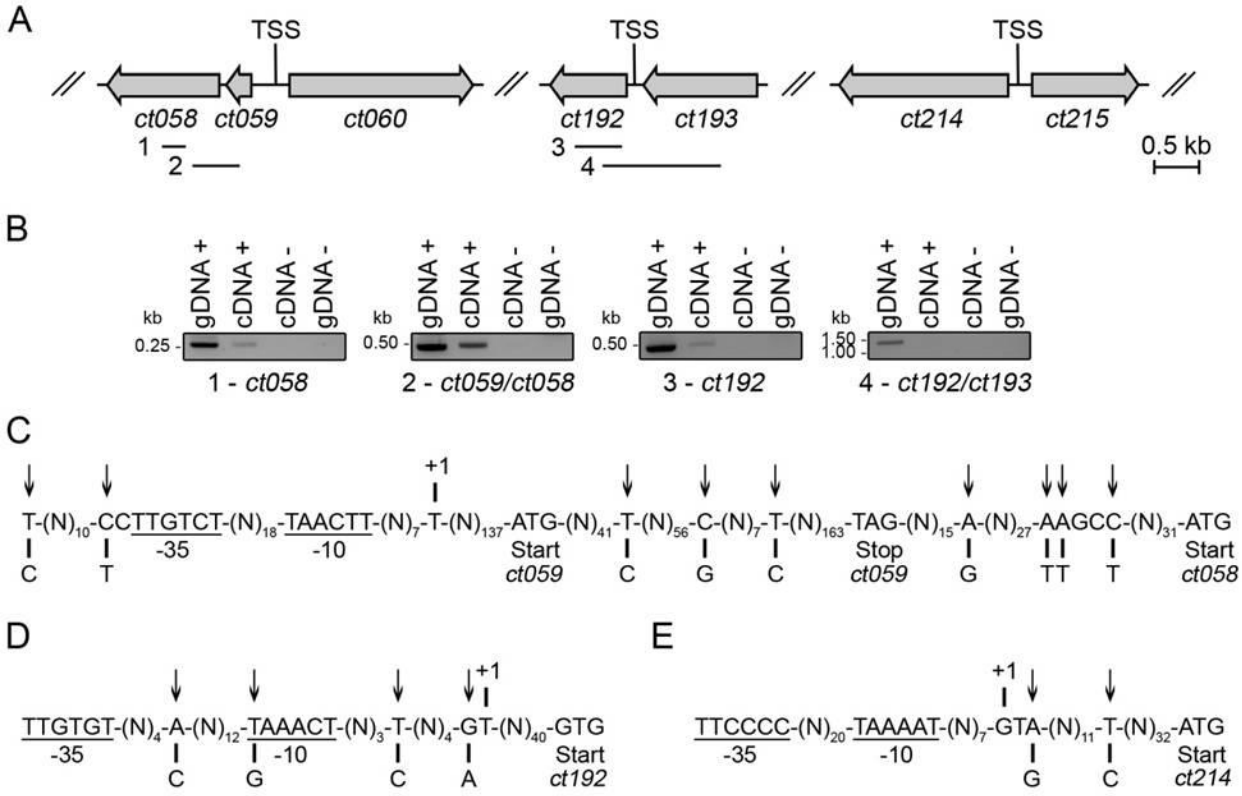


Figure 3.5. Identification of LGV-specific nucleotides in the promoter regions of *ct192* and *ct214* and within the *ct059-ct058* transcript. (A) Genetic organization of *ct058*, *ct192*, and *ct214* (the nomenclature of D/UW3 is used) depicting the fragments amplified in the transcriptional linkage analysis and the approximate localization of the transcriptional start sites (TSSs) determined by RACE in L2/434. (B) Transcriptional linkage analysis in L2/434: gDNA +, PCR from total DNA isolated from cells infected with L2/434; cDNA +, PCR from cDNA generated with reverse transcriptase (RT) from total RNA isolated from cells infected with L2/434; cDNA –, as cDNA+ but without RT; gDNA –, PCR from DNA of uninfected cells. (C to E) Schematic view of the nucleotide sequences of the *ct059* promoter region, *ct059-ct058* intragenic region, and first codons of *ct058* (C), and of the promoter regions of *ct192* (D) and *ct214* (E), in L2/434. The TSSs are labelled +1 and the predicted -10 and -35 σ^{66}-like hexamers are underlined. The sequences of the 3 LGV strains used in RT-qPCR assays (Figure 3.5) are 100% identical within the depicted regions (Supplemental Figures S3.3 and S3.4). The identified LGV-specific nucleotides are indicated with an arrow and below are the nucleotides present in those positions in the ocular and urogenital strains used in RT-qPCR assays (Figure 3.5). The full nucleotide sequences of these regions in the strains used in RT-qPCR assays are shown in Supplemental Figures S3.3 and S3.4.

C. trachomatis encodes three σ factors: σ^{66}, the homolog of *E. coli* main σ factor (σ^{70}); σ^{28}, a minor σ factor; and σ^{54}, an alternative σ factor (104). By inspecting the nucleotide sequences immediately upstream from the determined TSSs of *ct059-ct058*, *ct192*, and *ct214* for σ^{66}, σ^{54}, and σ^{28}-like promoters (104), we only identified σ^{66}-like promoters (Figure 3.5C-E, and Supplemental

Figures S3.3 and S3.4). Finally, we analyzed the promoter regions of *ct059-ct058*, *ct192*, and *ct214*, and the *ct059-ct058* transcript, for LGV-specific nucleotide differences, in all strains used in the RT-qPCR assays. We found LGV-specific nucleotides upstream from the predicted -35 region of *ct059* (Figure 3.5C and Supplemental Figure S3.3). Furthermore, we found 7 LGV specific nucleotides scattered within the coding region of *ct059* and in between the stop codon of *ct059* and the start codon of *ct058* (Figure 3.5C and Supplemental Figure S3.3). In particular, 3 of these LGV-specific nucleotide differences were clustered 30-35 nucleotides upstream from the start codon of *ct058* (Figure 3.5C and Supplemental Figure S3.3). However, it is unclear how these LGV-specific differences could explain the vestigial mRNA levels of *ct058* in ocular and urogenital strains or have a discriminatory role in the proposed hypothetical degradation of the *ct058* transcript in ocular and urogenital strains. The scenario was simpler for *ct192* and *ct214*, as we identified discrete LGV-specific nucleotides within the promoter regions of these genes that may explain their disease group-specific expression (Figure 3.5D and E, and Supplemental Figure S3.4).

3.5. Discussion

We found that amino acid differences between Inc proteins and distinct mRNA levels among *inc* genes throughout the developmental cycle of *C. trachomatis* strains correlate with the specific invasiveness and tropism of LGV isolates. Thus, we propose the novel hypothesis that a subset of Inc proteins may contribute for the specificity of infection by LGV strains. In fact, the vast majority of amino acid differences between Inc proteins are due to variations between proteins from LGV and trachoma biovars (Figure 3.2). This could simply reflect the evolutionary history of *C. trachomatis* (37). However, most *inc* genes showed d_N/d_S values of > 1 among *C. trachomatis* strains and, according to the Z-test of selection, 10 *inc* genes are likely under positive selection. In contrast, *pmps* or selected housekeeping genes of *C. trachomatis* mostly showed d_N/d_S values of < 1 (Figure 3.3). Moreover, polymorphisms in *C. trachomatis* genomes are essentially driven by fixation of silent mutations (301). This suggests that the amino acid differences between Inc proteins should not be solely explained by genetic drift. In addition, almost all disease group-specific amino acids of Inc proteins encoded by genes likely under positive selection were found among proteins of LGV strains, and the majority of these amino acids localized in regions of the proteins predicted to face the host cell cytosol (Table 3.2). Yet, it must be clarified that other proteins in addition to Inc proteins are likely involved in the specificity of infection by LGV strains (174,301). Moreover, the overall determinants of tissue tropism of *C. trachomatis* should be complex and multifactorial, and certainly also include e.g. the products of the *trpRBA* operon (51,180,285) or of the cytotoxin gene (179), and Tarp (287).

We have done a focused and in-depth study of the variability of Inc proteins and evolution and expression of *inc* genes among *C. trachomatis* serovars, encompassing almost all genomic information currently available. We reveal that the the overall degree of variation in the amino acid sequences of Inc proteins among strains is similar to that of a characteristic family of *C. trachomatis* polymorphic proteins (Pmps). Our results strengthen previous studies that suggested that some *C. trachomatis* Inc proteins could contribute to tissue tropism (163,292,301,302), and confirm recent data

suggesting that many *inc* genes could be under positive selection (174,301). However, almost all of these previous studies analyzed a limited number of sequences, as the majority of the 51 genomic sequences used in our work only became available very recently (37).

Our findings also revealed that differential gene expression could be a mechanism contributing to the different invasiveness and tissue tropism of *C. trachomatis* strains. A previous work have identified differences in expression between *pmp* genes from reference and clinical strains (277), but disease group-specific differences in expression of *C. trachomatis* genes have not been noticed before. The *inc* genes *ct058*, *ct192*, and *ct214* evidenced LGV-specific gene expression (Figure 3.4), and we have further identified LGV-specific nucleotides in the promoter regions of *ct192* and *ct214* and within the *ct059-ct058* transcript (Figure 3.5). This was directly analyzed for the strains used in RT-qPCR assays, but the specificity is maintained within all 51 *C. trachomatis* genomes (Supplemental Table S3.1) that we used in our studies (data not shown). As only 3 out of 48 *inc* genes showed LGV-specific gene expression, it is unlikely that this specificity is a common feature among *C. trachomatis* genes.

For *ct214*, we tentatively propose that the LGV-specific nucleotides could differentially affect its expression either at transcriptional or post-transcriptional level. For example, a recent study showed that a *Salmonella* small non-coding RNA can discriminate mRNA regions that differ in one single nucleotide (303). The picture is more complex for *ct058* and *ct192*. Regarding *ct058*, in L2/434 this gene is co-transcribed with *ct059* and mRNA levels of *ct059* were previously detected also in ocular and urogenital strains (231), which we re-confirmed (data not shown). We speculate that specific processing of the *ct059-ct058* transcript selectively reduces the levels of *ct058* mRNA in ocular and urogenital strains. Furthermore, *ct058* is a pseudogene in many ocular strains (but not in those used in the RT-qPCR assays; Supplemental Table S3.3 and data not shown) and *ct192* is a pseudogene in at least ocular strain B/Jali20 (Supplemental Table S3.3). This does not necessarily have an impact in the interpretation of the data, as for example the LGV-specific pseudogene *ct300* showed similar mRNA levels during the developmental cycle of C/TW3, E/Bour or L2/434 (Supplemental Table S3.6). Another issue is that there are different annotations in GenBank of the start codon of *ct192* in the archetype ocular (A/Har13), urogenital (D/UW3) and LGV (L2/434) strains (Supplemental Table S3.3). This alone could explain the differential expression of *ct192*, owing to dissimilar promoter regions. However, inspection of the nucleotide sequence immediately upstream from the annotated start codons of *ct192* in each of these three strains only reveals a strong putative ribosome binding site in L2/434 (Supplemental Table S3.3). Therefore, it is likely that the start codon of *ct192* is conserved between *C. trachomatis* strains and corresponds to the one annotated in L2/434. In this situation, *ct192* is a pseudogene in all ocular strains (Supplemental Table S3.3) and the differences in the mRNA levels of *ct192* between LGV and ocular or urogenital strains may be tentatively explained by the single nucleotide differences found within its promoter region (Figure 3.5D, and Supplemental Figure S3.4..

It is not possible to make a rigorous side-by-side comparison between our RT-qPCR data and previous analyses of *inc* gene expression by RT-PCR (98) and microarrays (99,100), as the sensitivity of the methods used is quite different. In general, we have confirmed that *inc* genes display different

profiles of expression but that they are mostly early-cycle genes, which supports the idea that many Inc proteins should play a role in modifying the inclusion membrane early during development (103,127). We also found *inc* genes displaying an early- and late-cycle profile of expression. A similar profile has been observed in microarray studies (99), and it was suggested to result from “carry-over” mRNA of highly expressed late genes from the previous infectious cycle. However, the first time-point we analyzed was at 2 h post-infection, when “carry-over” mRNA would have been degraded. Furthermore, at 30 or 42 h post-infection, several late-cycle *inc* genes showed levels of expression at least comparable to those of early- and late-cycle *inc* genes (Supplemental Table S3.6). This suggests that the expression of some *inc* genes could be induced early in the cycle, down-regulated at mid-cycle, and induced again at late-cycle.

Not all predicted *C. trachomatis* Inc proteins have been detected on the inclusion membrane. We addressed this by the analysis of T3S signals in 48 *C. trachomatis* Inc proteins. This led to the identification of 19 novel T3S substrates, including 14 putative Inc proteins (Figure 3.1 and Table 3.1). In particular, and importantly, putative Inc proteins CT058, CT192, and CT214 showed a T3S signal. Furthermore, all 10 *inc* genes likely under positive selection encode known Inc proteins. Our data for T3S signals in 14 Inc proteins that were previously analyzed using *Shigella flexneri* as heterologous bacteria (129,132) (Table 3.1) only differed for CT192 and CT850. This could be related to the distinct heterologous hosts used, or to the analysis of different N-terminal regions of CT192 (due to the different annotations of the start codon of *ct192* in GenBank).

Although human macrophages have strong antimicrobial activity against *C. trachomatis* ocular and urogenital strains, they support the growth of LGV strains (304). We hypothesize that specific amino acids in Inc proteins, or their earlier and/or higher expression, could specifically enable LGV strains to inhibit phagolysosomal fusion in macrophages and/or prevent the formation of reactive oxygen or nitrogen species (305). Unfortunately, little is known about the function of the Inc proteins that we have identified as potentially involved in these processes. CT222 and CT118/IncG co-localize with kinases of the Src family in discrete regions of the inclusion membrane that associate with host cell centrosomes (258), and CT223 could inhibit host cell cytokenesis (260). However, it is unclear how this may relate to our hypothesis. On the other hand, Inc proteins that could manipulate intracellular membrane trafficking, such as those that have SNARE-like motifs (CT119/IncA, CT223, and CT813) (255) or interact with Rab GTPases (CT229) (294), are good candidates to selectively inhibit macrophage phagolysosomal fusion. Not all *C. trachomatis* Inc proteins tentatively proposed to be specifically involved in inhibiting phagolysosomal fusion have homologues in other chlamydial species, which also avoid this host cell degradation pathway (129,292). It is possible that each *Chlamydia*l species evolved particular Inc proteins, and other virulence proteins, to account for the specificity of each type of cell it infects.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through grants PEst-OE/EQB/LA0004/2011 and PTDC/SAU-MII/099623/2008, and by the European Commission through a Marie Curie European Re-integration Grant (PERG03-GA-2008-230954) to LJM. FA, VB, and RF held PhD fellowships SFRH/BD/73545/2010, SFRH/BD/68527/2010, and SFRH/BD/68532/2010, respectively, from FCT. We thank Irina Franco for critical review of the manuscript.

Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system

This chapter includes the contents of the following publication, as well as additional data resulting from an extent evaluation of the expression profiles of the novel type III secretion substrates in C. trachomatis strains representing the three disease groups. The decision to include the remainder contents of the published manuscript that were not performed by the author of the present Ph.D. thesis relies on the need to put the data into the appropriate context.

da Cunha M., Milho C., Almeida F., Pais S.V., Borges V., Maurício R., Borrego M.J., Gomes J.P., Mota L.J.
Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system.
2014, BMC Microbiology, 14:40.
(http://www.ncbi.nlm.nih.gov/pubmed/24533538)

Personal contribution

VB carried out all the RT-qPCR assays (both the ones yielding the published data and the additional assays leading to the extra data that was included in this chapter), as well as the analysis of genetic variability within the promoter region of *ct105*.

4. Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system

4.1. Abstract

Background: *Chlamydia trachomatis* is an obligate intracellular human pathogen causing ocular and urogenital infections that are a significant clinical and public health concern. This bacterium uses a type III secretion (T3S) system to manipulate host cells, through the delivery of effector proteins into their cytosol, membranes, and nucleus. In this work, we aimed to find previously unidentified *C. trachomatis* T3S substrates.

Results: We first analyzed the genome of *C. trachomatis* L2/434 strain for genes encoding mostly uncharacterized proteins that did not appear to possess a signal of the general secretory pathway and which had not been previously experimentally shown to be T3S substrates. We selected several genes with these characteristics and analyzed T3S of the encoding proteins using *Yersinia enterocolitica* as a heterologous system. We identified 23 *C. trachomatis* proteins whose first 20 amino acids were sufficient to drive T3S of the mature form of β-lactamase TEM-1 by *Y. enterocolitica*. We found that 10 of these 23 proteins were also type III secreted in their full-length versions by *Y. enterocolitica*, providing additional support that they are T3S substrates. Seven of these 10 likely T3S substrates of *C. trachomatis* were delivered by *Y. enterocolitica* into host cells, further suggesting that they could be effectors. Finally, real-time quantitative PCR analysis of expression of genes encoding the 10 likely T3S substrates of *C. trachomatis* strains representing the three disease-groups (ocular, genital, and lymphogranuloma venereum) showed that they display distinct expression profiles throughout the developmental cycle, and that some of them potentially display disease-specific differential expression.

Conclusions: Using *Y. enterocolitica* as a heterologous system, we identified 10 likely T3S substrates of *C. trachomatis* (CT053, CT105, CT142, CT143, CT144, CT161, CT338, CT429, CT656, and CT849) and could detect translocation into host cells of CT053, CT105, CT142, CT143, CT161, CT338, and CT429. Therefore, we revealed several *C. trachomatis* proteins that could be effectors subverting host cell processes. Additionally, gene expression data suggest that the novel putative effectors are potentially involved in dissimilar mechanisms towards the host cell manipulation, and some of them may contribute to the distinct tropism of *C. trachomatis* strains.

4.2. Introduction

Chlamydiae are a large group of obligate intracellular bacteria that includes human pathogens (e.g. *Chlamydia trachomatis* or *C. pneumoniae*), animal pathogens (e.g. *C. abortus*, *C. caviae*, *C. felis*, or *C. muridarum*), or symbionts of free-living amoebae. Among *Chlamydiae*, *C. trachomatis* is a particular clinical and public health concern, being the leading cause of infectious blindness in developing countries (6) and the most prevalent sexually transmitted bacteria worldwide (4).

Like all *Chlamydiae*, *C. trachomatis* undergoes a developmental cycle involving the inter-conversion between two morphologically distinct forms: a non-replicative infectious form, the elementary body (EB), and a replicative non-infectious form, the reticulate body (RB) (88). Throughout its developmental cycle, *C. trachomatis* uses a type III secretion system (T3SS) to translocate several effector proteins across the host cell plasma membrane and the inclusion membrane (103,127). These T3S effectors are thought to play a central role in bacterial invasion (141,306) and exit of host cells (307), and in the subversion of various host cell processes (255,258,266,290,293,294,308,309). There are, however, chlamydial effectors, such as CPAF/CT858 or CT441, which are not T3S substrates (127).

Given their likely central role during infection, considerable efforts have been placed at identifying chlamydial effectors. This is not a trivial task because the amino acid sequence of most effectors does not display significant similarity to proteins of known function. Additionally, T3S substrates, which should comprise the bulk of *Chlamydia* effectors, contain no easily recognizable secretion signal. Moreover, in spite of the recent development of systems for transformation of *Chlamydia* (206,215), for a long time no methods have been available for genetic manipulation of these bacteria. To overcome these obstacles, chlamydial effectors have been searched: *i*) by systematic phenotypic analyses of yeast *Saccharomyces cerevisiae* expressing individual chlamydial proteins (156); *ii*) by using *Salmonella* (288), *Shigella* (132,309-311), or *Yersinia* (140,266,290,291,312,313) as genetically tractable heterologous host bacteria carrying well characterized T3SSs; or *iii*) by complex computational predictions of T3S signals (314-316). The subsequent use of specific antibodies enabled to detect translocation into host cells of some of the *C. trachomatis* proteins singled out in these searches, such as in the case of Tarp/CT456 (140), CT694 (266), CopN/CT089 (291), Cap1/CT529 (160), CT620 (310), CT621 (310,317), CT711 (310), lipid-droplet associated (Lda) proteins Lda1/CT156, Lda2/CT163, and Lda3/CT473 (318), Nue/CT737 (309), or of a group of proteins containing a hydrophobic motif thought to mediate their insertion into the inclusion membrane (Inc proteins) (258,300). Moreover, the direct use of antibodies raised against particular *C. trachomatis* proteins (CT311, CT622, CT795, GlgA/CT798, HtrA/CT823, or Pgp3) revealed their presence in the host cell cytosol or nucleus of infected cells (265,319-323). Finally, the *in vitro* deubiquitinase activity of *Chla*DUB1/CT868 and of *Chla*DUB2/CT867 (268), and the capacity of *Chla*DUB1/CT868 to suppress the NF-κB pathway in transfected cells (269), indicate that these two proteins should be effectors.

In this work, we have surveyed the genome of *C. trachomatis* mostly for genes encoding uncharacterized proteins that were not described before as T3S substrates. We then used *Yersinia enterocolitica* as a heterologous system to identify 10 novel likely T3S substrates of *C. trachomatis* and real-time quantitative PCR (RT-qPCR) to show that 9 of the genes encoding these proteins are clearly expressed during the bacterial developmental cycle. Furthermore, we showed that 7 of the 10 likely T3S substrates of *C. trachomatis* could be translocated into host cells by *Y. enterocolitica*. Therefore, we identified several novel putative effectors of *C. trachomatis*.

4.3. Material and Methods

4.3.1. Cell culture, bacterial strains and growth conditions

HeLa 229 (ATCC) cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Invitrogen) supplemented with 10% (v/v) foetal bovine serum (FBS; Invitrogen) at 37^{o}C in a humidified atmosphere of 5% (v/v) CO_2. *C. trachomatis* L2/434/Bu (from ATCC) was propagated in HeLa 229 cells using standard techniques (193). *Escherichia coli* TOP10 (Invitrogen) was used for construction and purification of the plasmids. *Yersinia enterocolitica* ΔHOPEMT (MRS40 pIML421 [$yopH_{\Delta 1\text{-}352}$, $yopO_{\Delta 65\text{-}558}$, $yopP_{23}$, $yopE_{21}$, $yopM_{23}$, $yopT_{135}$]), deficient for the *Yersinia* Yop T3S effectors H, O, P, E, M, and T, but T3S-proficient (295) and T3S-deficient *Y. enterocolitica* ΔHOPEMT ΔYscU (MRS40 pFA1001 [$yopH_{\Delta 1\text{-}352}$, $yopO_{\Delta 65\text{-}558}$, $yopP_{23}$, $yopE_{21}$, $yopM_{23}$, $yopT_{135}$, $yscU_{\Delta 1\text{-}354}$) (324) were used for T3S assays. The *yscU* gene encodes an essential component of the *Y. enterocolitica* T3S system, and the $yscU_{\Delta 1\text{-}354}$ mutation is non-polar (296). *E. coli* or *Y. enterocolitica* were routinely grown in liquid or solid Luria-Bertani (LB) medium with the appropriate antibiotics and supplements. Plasmids were introduced into *E. coli* or *Y. enterocolitica* by electroporation.

4.3.2. DNA manipulations, plasmids, and primers

The plasmids used in this work and their main characteristics are detailed in Supplemental Table S4.1. The DNA primers used in their construction are shown Supplemental Table S4.2. Plasmids were constructed and purified with proof-reading Phusion DNA polymerase (Finnzymes), restriction enzymes (MBI Fermentas), T4 DNA Ligase (Invitrogen), DreamTaq DNA polymerase (MBI Fermentas), DNA clean & concentrator^{TM}-5 Kit and Zymoclean^{TM} Gel DNA Recovery kit (Zymo Research), and purified with GeneElute Plasmid Miniprep kit (Sigma), according to the instructions of the manufacturers. In brief, to analyze T3S signals we constructed plasmids harboring hybrid genes encoding the first 10, 15, 20, or 40 amino acids of each protein *(C. trachomatis* proteins, SycT and YopE) and the mature form of TEM-1 β-lactamase (TEM-1) (297). These hybrids were made using as vector plasmid pLJM3, a low-copy plasmid which enables expression of the cloned genes driven by the promoter of the *Y. enterocolitica yopE* gene (298), either by overlapping PCR or by using a cloning strategy previously described for the construction of plasmids encoding Inc-TEM-1 hybrid proteins (324). To analyze secretion of full-length *C. trachomatis* proteins, we constructed plasmids expressing the proteins C-terminally tagged with a haemagglutinin (HA) epitope. For this, the genes were amplified by PCR from chromosomal DNA of strain L2/434/Bu using a reverse primer with a sequence complementary to the transcribed strand of the DNA encoding the HA-epitope. PCR products digested with the appropriate enzymes were ligated into pLJM3 (298). The accuracy of the nucleotide sequence of all the inserts in the constructed plasmids was checked by DNA sequencing.

4.3.3. *Y. enterocolitica* T3S assays

T3S assays were done as previously described (296). We used *Y. enterocolitica* ΔHOPEMT or ΔHOPEMT ΔYscU strains carrying the plasmids described in Supplemental Table S4.1. The proteins in bacterial pellets and culture supernatants were analyzed by immunoblotting, and the amounts of protein in bacterial pellets and/or culture supernatants were estimated from images of immunoblots with Image Lab (Bio-Rad). Where appropriate, we calculated the percentage of secretion as the ratio between the amounts of secreted protein (in the culture supernatant fraction) relative to the total amount of protein (in the culture supernatant and in the bacterial pellet fractions). The results from the quantifications are the average ± standard error of the mean (SEM) from at least three independent experiments. Detailed results for each protein analyzed are in Supplemental Table S4.3.

4.3.4. *Y. enterocolitica* translocation assays

Analyses of protein translocation into host cells by *Y. enterocolitica* were done essentially as previously described (325,326). In brief, *Y. enterocolitica* strains were grown in brain heart infusion (BHI; Scharlau) medium overnight at 26^{o}C with continuous shaking (130 rpm). Bacteria were then diluted to an optical density at 600 nm of 0.2 in fresh BHI and cultured in the same conditions for 2 h. Subsequently, the *yop* regulon was induced by incubation for 30 min in a shaking water bath (130 rpm) at 37^{o}C. Bacteria were then washed with DMEM supplemented with 10% (v/v) FBS and added to HeLa 229 cells, grown overnight in 24-well plates (1x10^{5} cells/well), by using a multiplicity of infection of 50. The infected cells were incubated at 37^{o}C in a humidified atmosphere of 5% (v/v) CO_2. After 3 h of incubation, extracellular bacteria were killed by adding gentamicin (50 µg/ml), and the cells were incubated in the same conditions for additional 2 h. The infected cells were then harvested on ice, washed with phosphate-buffered saline (PBS), ressuspended in PBS containing 0.1% (v/v) Triton X-100 and a protease inhibitor cocktail (Sigma), and incubated for 10 min on ice. The samples were then centrifuged (15,000 *g* for 15 min at 4^{o}C) and Triton-soluble and Triton-insoluble HeLa cell lysates were loaded on sodium dodecyl sulfate-12% (v/v) polyacrilamide gels. After electrophoresis, the gels were processed for immunoblotting using 0.2 µm pore-size nitrocellulose membranes (BioRad).

4.3.5. Immunoblotting

The following antibodies were used for immunoblotting: rat monoclonal anti-HA (3F10; Roche; 1:1000), mouse monoclonal anti-TEM-1 (QED Bioscience; 1:500), rabbit polyclonal anti-SycO (1:1000) (299), and mouse monoclonal anti-tubulin (clone B-5-1-2; Sigma; 1:1000). Immunoblot detection was done with horseradish peroxidase-conjugated secondary antibodies (GE Healthcare and Jackson ImmunoResearch), Western Lightning *Plus*-ECL (Perkin Elmer), and a ChemiDoc XRS+ system (BioRad) or exposure to Amersham Hyperfilm ECL (GE Healthcare). All quantitative analyses were done with immunoblot images obtained using ChemiDoc XRS+ (BioRad).

4.3.6. Real-time quantitative PCR

The expression of newly identified candidate T3S substrates during the developmental cycle of *C. trachomatis* L2/434, C/TW-3 and E/Bour was estimated by determining mRNA levels at different times post-infection by real-time quantitative PCR (RT-qPCR). These experiments were done as previously described (324). Primers (available upon request) were designed using Primer Express (Applied Biosystems). The RT-qPCR assays were done using the ABI 7000 SDS, SYBR green chemistry, and optical plates (Applied Biosystems), as previously described (231). At each time point, raw RT-qPCR data for each gene were normalized against the data obtained for the 16S rRNA transcript, as it was previously demonstrated that this is an adequate endogenous control (231). The final results were based on three independent experiments.

4.4. Results

4.4.1. Selection of *C. trachomatis* proteins analyzed in this work

To search for previously unidentified T3S substrates of *C. trachomatis*, we first surveyed the genome of strain L2/434 for genes encoding mostly uncharacterized proteins, or with a putative biochemical activity compatible with the function of a T3S effector (e.g., proteases). Among these genes, we selected those encoding proteins not predicted to have a signal sequence characteristic of the general secretory pathway (according to Psortb v3.0) and that had not been previously analyzed experimentally for the presence of a T3S signal. This singled out 32 proteins (CT016, CT017, CT031, CT051, CT053, CT080, CT105, CT142, CT143, CT144, CT153, CT161, CT172, CT273, CT277, CT289, CT309, CT330, CT338, CT386, CT425, CT568, CT583, CT590, CT631, CT635, CT656, CT696, CT702, CT837, CT845, and CT849; we used the nomenclature of the annotated *C. trachomatis* D/UW3 strain (115); the names of the corresponding genes as annotated for strain L2/434 (46) can be found Supplemental Table S4.3). Furthermore, for comparison purposes, we considered proteins that had been tested for the presence of a T3S signal using *Shigella flexneri* as a heterologous bacteria: eight proteins whose first ~40 amino acids of the corresponding *C. pneumoniae* homologs did not drive secretion of an adenylate cyclase (Cya) reporter protein by *S. flexneri* (CT066, CT429, GrgA/CT504, CT538, CT584, CT768, CT779, CT814), and three proteins whose N-terminal region of the *C. pneumoniae* homologs drove secretion of a Cya reporter protein by *S. flexneri* (CT203, CT577, CT863) (132). Please note that at the time this work was initiated GrgA/CT504 was an uncharacterized protein; however, it was recently described as a transcriptional activator (327). Finally, throughout this study we used as positive controls a *C. trachomatis bona-fide* T3S effector (CT694) (266) and a *C. trachomatis* likely T3S substrate (CT082) that we had previously identified (312), and which was recently independently confirmed (313), and as negative control a predicted ribosomal protein (RplJ/CT317).

In summary, in experiments that will be described below, we analyzed T3S signals in 46 *C. trachomatis* proteins (~5% of all proteins encoded by the L2/434 strain): 32 hypothetical proteins

previously not analyzed experimentally for T3S signals, 11 proteins whose *C. pneumoniae* homologs were previously analyzed for T3S signals using *S. flexneri* as heterologous system, and 3 controls. In the selection of these proteins, we did not consider predictions made by any of the published *in silico* methods that suggest putative T3S substrates (314-316,328).

4.4.2. The first 20 amino acids of *C. trachomatis* T3S substrates are sufficient to drive efficient secretion of TEM-1 hybrid proteins by *Y. enterocolitica*

We previously used TEM-1 as a reporter protein to analyze T3S signals in *C. trachomatis* Inc proteins, using *Y. enterocolitica* as a heterologous system (324). However, before analyzing T3S signals in the proteins that we selected to study in this work (see above), we sought to ascertain the optimal amino acid length of the chlamydial T3S signal that drives secretion of TEM-1 hybrid proteins in *Yersinia*. For this, we analyzed secretion of hybrid proteins comprising the first 10, 20 and 40 amino acids of known *C. trachomatis* T3S substrates (IncA or IncC) fused to TEM-1 ($IncA_{10}$-TEM-1, $IncA_{20}$-TEM-1, $IncA_{40}$-TEM-1, $IncC_{10}$-TEM-1, $IncC_{20}$-TEM-1, $IncC_{40}$-TEM-1) by T3S-proficient (ΔHOPEMT) or T3S-deficient (ΔHOPEMT ΔYscU) *Y. enterocolitica* (Figure 4.1). As negative controls we analyzed secretion by *Y. enterocolitica* ΔHOPEMT of TEM-1 alone and of a hybrid protein comprising the first 20 amino acids of the *Yersinia* T3S chaperone SycT to TEM-1 ($SycT_{20}$-TEM-1), and as positive control we analyzed secretion by ΔHOPEMT of a fusion of the first 15 amino acids of the *Yersinia* effector YopE to TEM-1 ($YopE_{15}$-TEM-1) (Figure 4.1), an archetypal T3S signal (329,330). Bacteria expressing these proteins were incubated under T3S-inducing conditions, as described in Methods. As expected, and in agreement to what we previously reported (324), mature TEM-1 alone was not secreted and the $SycT_{20}$-TEM-1 fusion showed a percentage of secretion of 3.0 (SEM, 0.3). Based on this, to decide if a TEM-1 hybrid was secreted or not we set the threshold of percentage of secretion to 5 (Figure 4.1A). The six Inc-TEM-1 hybrid proteins were type III secreted (Figure 4.1A and Figure 4.1B). However, $IncA_{10}$-TEM-1 and $IncC_{10}$-TEM-1 were secreted less efficiently than $YopE_{15}$-TEM-1, while $IncA_{20}$-TEM-1, $IncA_{40}$-TEM-1, $IncC_{20}$-TEM-1 and $IncC_{40}$-TEM-1 were secreted at levels comparable to $YopE_{15}$-TEM-1 (Figure 4.1A). Overall, these experiments indicated that the first 20 amino acids of *C. trachomatis* T3S substrates are sufficient to drive secretion of TEM-1 hybrid proteins by *Y. enterocolitica* ΔHOPEMT as efficiently as the first 15 amino acids of the *Yersinia* effector YopE.

4.4.3. Identification of T3S signals in *C. trachomatis* proteins

To identify T3S signals in the selected 46 *C. trachomatis* proteins, we analyzed secretion of fusions to TEM-1 of the first 20 amino acids of each of these proteins by T3S-proficient *Y. enterocolitica* ΔHOPEMT. These experiments revealed 24 *C. trachomatis* proteins whose first 20 amino acids drove secretion of TEM-1 hybrid proteins by *Y. enterocolitica* (Figure 4.2A). Owing to lack of expression, or very low expression levels, it was not possible to conclude if the TEM-1 hybrids comprising the N-terminal region of CT590, CT845 and CT863 were secreted (Figure 4.2A).

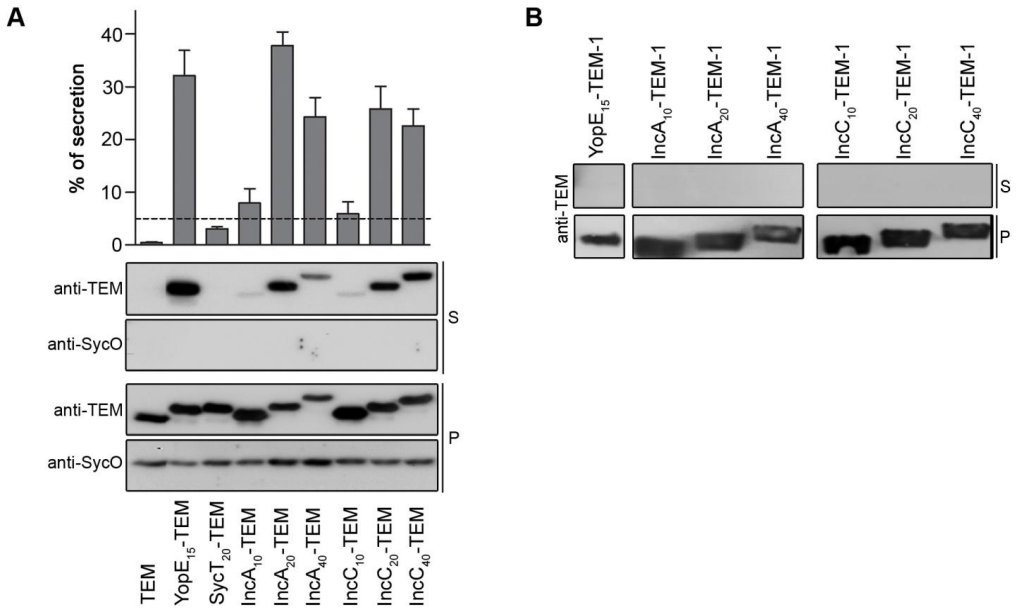


Figure 4.1. The first 20 amino acids of known *C. trachomatis* T3S substrates (IncA or IncC) are sufficient to efficiently drive T3S of TEM-1 hybrid proteins by *Y. enterocolitica*. *Y. enterocolitica* T3S-proficient (ΔHOPEMT) (A) and T3S-defective (ΔHOPEMT ΔYscU) (B) were used to analyze secretion of hybrid proteins comprising the first 10, 20, or 40 amino acids of *C. trachomatis* IncA or IncC, or the first 15 or 20 amino acids of *Y. enterocolitica* YopE or SycT, respectively, fused to the mature form of TEM-1 β-lactamase (TEM-1). Immunoblots show the result of T3S assays in which proteins in culture supernatants (S, secreted proteins) and in bacterial pellets (P, nonsecreted proteins) from ~$5x10^7$ bacteria were loaded per lane. The first 15 amino acids of the *Yersinia* effector YopE correspond to an archetypal T3S signal (329,330), and $YopE_{15}$-TEM-1 was used as positive control; SycT and SycO are strictly cytosolic *Yersinia* T3S chaperones (295,299). $SycT_{20}$-TEM-1 was a negative control for the T3S assays. Immunodetection of SycO ensured that the presence of TEM-1 hybrid proteins in the culture supernatants was not a result of bacterial lysis or contamination. The percentage (%) of secretion of each TEM-1 hybrid was calculated by densitometry, as the ratio between the amount of secreted and total protein. The threshold to decide whether a protein was secreted was set to 5% (dashed line), based on the % of secretion of $SycT_{20}$-TEM-1. Data are the mean ± SEM from at least 3 independent experiments.

By individually introducing the plasmids encoding the TEM-1 hybrid proteins that were secreted into T3S-deficient *Y. enterocolitica* ΔHOPEMT ΔYscU and performing T3S assays, we confirmed that secretion of the proteins was dependent on a functional T3SS (Figure 4.2B). The percentage of secretion of the different hybrid proteins that were secreted varied considerable, between 56% (SEM, 4) for $CT694_{20}$-TEM-1 to 5% (SEM, 2) for $CT143_{20}$-TEM-1 (Figure 4.2B). Overall, this confirmed a T3S signal in CT203, which has been previously shown to be a T3S substrate (132), and revealed T3S signals in 23 previously unrecognized T3S substrates of *C. trachomatis*.

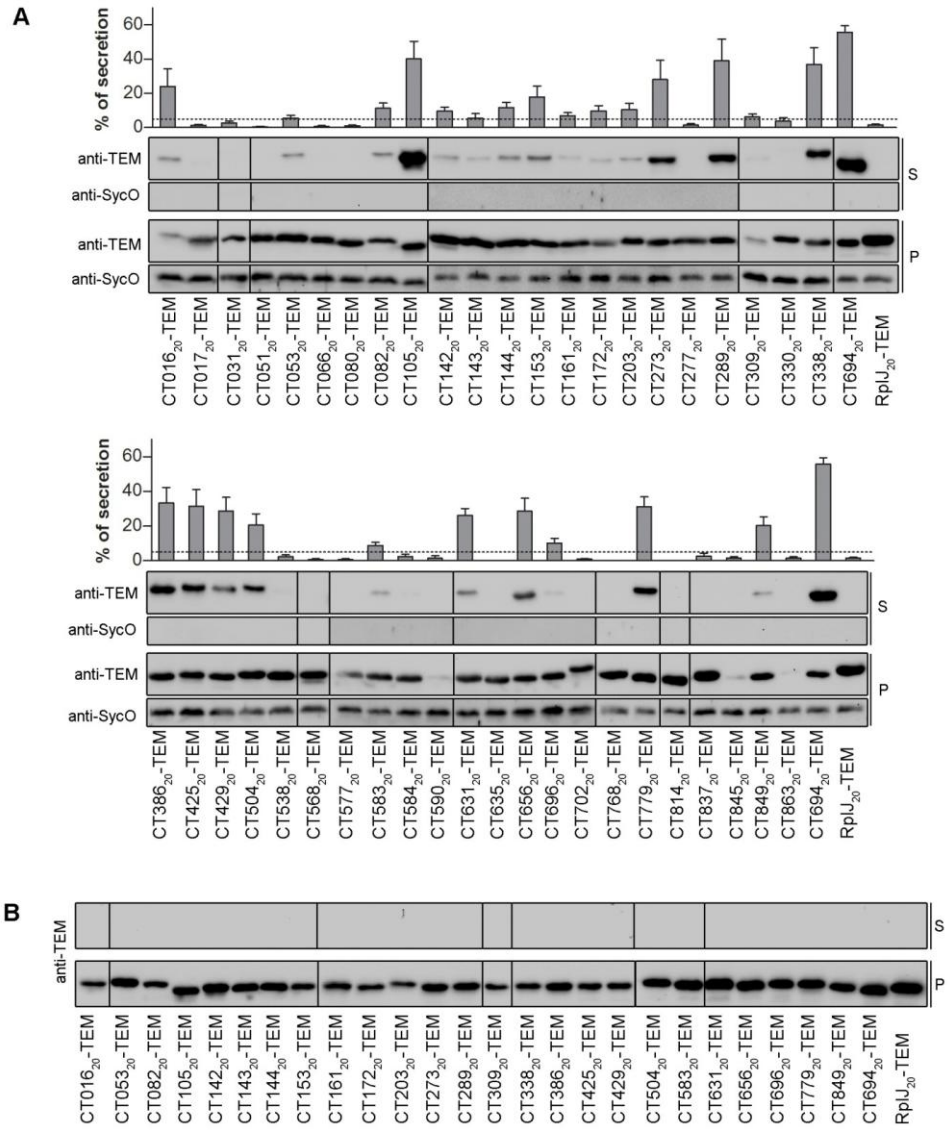


Figure 4.2. Identification of T3S signals in *C. trachomatis* proteins using *Y. enterocolitica* as a heterologous system. *Y. enterocolitica* T3S-proficient (ΔHOPEMT) (A) and T3S-defective (ΔHOPEMT ΔYscU) (B) were used to analyze secretion of hybrid proteins comprising the first 20 amino acids of selected *C. trachomatis* proteins or the first 20 amino acids of *Y. enterocolitica* SycT fused to the mature form of TEM-1 β-lactamase (TEM-1). Immunoblots show the result of T3S assays in which proteins in culture supernatants (S, secreted proteins) and in bacterial pellets (P, non-secreted proteins) from ~2.5x10^8 and ~5x10^7 bacteria, respectively, were loaded per lane. TEM-1 hybrids of the known *C. trachomatis* T3S substrates CT082 (312,313) and CT694 (266) were used as positive controls. SycT and SycO are strictly cytosolic *Yersinia* T3S chaperones (295,299). SycT$_{20}$-TEM-1 was a negative control for the T3S assays. Immunodetection of SycO ensured that the presence of TEM-1 hybrid proteins in the culture supernatants was not a result of bacterial lysis or contamination. The percentage (%) of secretion of each TEM-1 hybrid was calculated by densitometry, as the ratio between the amount of secreted and total protein. The threshold to decide whether a protein was secreted was set to 5% (dashed line), based on the % of secretion of SycT$_{20}$-TEM-1. Data are the mean ± SEM from at least 3 independent experiments.

4.4.4. Analysis of the secretion of the newly identified candidate T3S substrates of *C. trachomatis* as full-length proteins

We next analyzed if the 23 *C. trachomatis* proteins carrying newly identified T3S signals, and also CT203 and the controls (CT082, CT694 and RplJ), were secreted as full-length proteins by *Y. enterocolitica* ΔHOPEMT. The rationale for these experiments was that some proteins cannot be type III secreted even with a T3S signal grafted at their N-termini (331-334), possibly because the secretion channel is too narrow (inner diameter of 2-3 nm (139)) to accommodate tightly folded proteins. For example, while we showed that $YopE_{15}$-TEM-1 is efficiently type III secreted, hybrid proteins containing the first 15 or 16 amino acids of YopE fused to mouse dihydrofolate reductase (DHFR) are not type III secreted by *Y. enterocolitica* (331,332). This indicates that most T3S substrates must have particular folding properties that are compatible with them being type III secreted proteins. Based on this, we predicted that if the full-length version of chlamydial proteins were type III secreted by *Yersinia* this would be an additional indication that they can be T3S substrates. However, lack of secretion of the full-length proteins would not preclude that they could be T3S substrates, as they may require *Chlamydia*-specific chaperones, not present in *Yersinia* (335).

To analyze secretion of full-length *C. trachomatis* proteins by *Y. enterocolitica* we used plasmids expressing the chlamydial proteins with an HA tag at their C-termini. The plasmids were introduced into *Y. enterocolitica* ΔHOPEMT and T3S assays were performed. In these experiments, the percentage of secretion of the positive controls (CT694-HA and CT082-HA) was between 20-30% and the percentage of secretion of the negative control (RplJ-HA) was 0.13% (SEM, 0.05). Based on these results, in experiments involving full-length proteins of newly identified chlamydial T3S substrates we set a conservative threshold of 2% to decide whether a protein was secreted or not. This defined a group of 11 proteins that in their full-length version were secreted by *Y. enterocolitica* ΔHOPEMT: CT053-HA, CT105-HA, CT142-HA, CT143-HA, CT144-HA, CT161-HA, CT338-HA, CT429-HA, CT583-HA, CT656-HA, and CT849-HA (Figure 4.3A and 4.3B). To test if secretion of these proteins was dependent on a functional T3SS, the plasmids carrying their encoding genes, as well as plasmids encoding positive controls CT694-HA or CT082-HA, were individually introduced into T3S-deficient *Y. enterocolitica* ΔHOPEMT ΔYscU. With the exception of CT583-HA, which for unknown reasons was very poorly expressed by *Y. enterocolitica* ΔHOPEMT ΔYscU, these assays indicated that the other 10 proteins analyzed were type III secreted (Fig. 3C).

Secretion of full-length CT153-HA, CT172-HA, CT203-HA, CT386-HA or CT425-HA by *Y. enterocolitica* could occasionally be seen by immunoblotting (Fig. 3A); however, the results were not always reproducible and their individual average percentage of secretion was in all cases below 2% (Fig. 3B). We did not detect significant amounts of CT273-HA, CT289-HA, CT309-HA, or CT631-HA in culture supernatants (Figure 4.3A and Supplemental Table S4.3), but as their levels of expression were either extremely low (CT273-HA, CT289-HA, and CT309-HA) or undetectable (CT631-HA) it was not possible to draw conclusions about secretion of these proteins. Furthermore, CT016-HA, and possibly CT696-HA (barely visible in Fig. 3A), were immunodetected in the culture supernatant fraction in a form that migrated on SDS-PAGE at a molecular weight much lower than the one predicted from

their amino acid sequence (27 kDa and 46 kDa, respectively), while in the bacterial pellet fraction their migration on SDS-PAGE corresponded roughly to their predicted molecular weight (Figure 4.3A). This suggests that the proteins could be cleaved during secretion, unstable in the culture supernatant, or their encoding genes possess internal Shine-Dalgarno sequences. Regardless of the exact reason, we could not confidently analyze whether CT016-HA and CT696-HA were secreted or not.

Overall, the full set of T3S assays revealed 10 proteins (CT053, CT105, CT142, CT143, CT144, CT161, CT338, CT429, CT656, and CT849) as newly identified likely T3S substrates of *C. trachomatis*, and therefore as possible effectors.

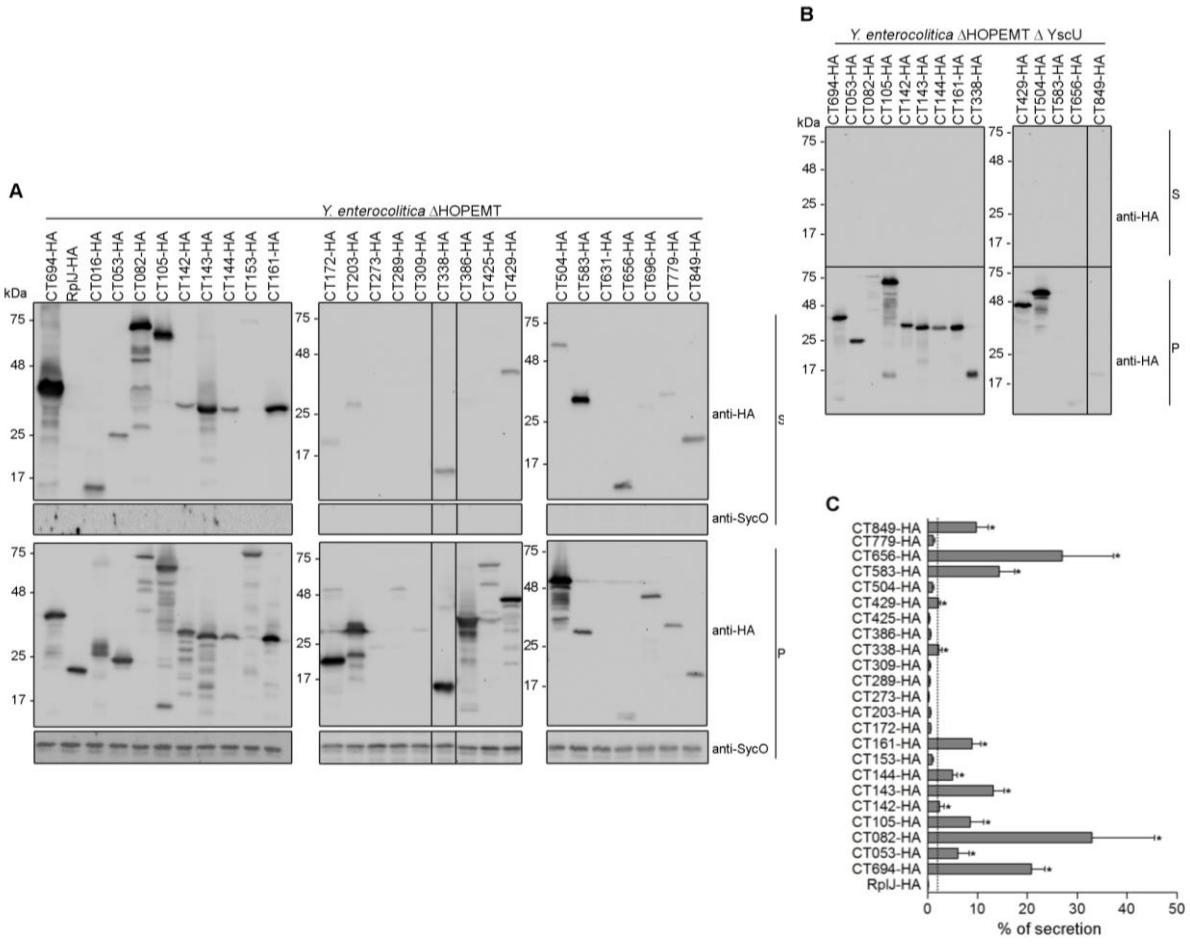


Figure 4.3. Analysis of the T3S of *C. trachomatis* full-length proteins by *Y. enterocolitica*. *Y. enterocolitica* T3S-proficient (ΔHOPEMT) (A) and T3S-defective (ΔHOPEMT ΔYscU) (B) were used to analyze secretion of full-length *C. trachomatis* proteins with a C-terminal HA epitope tag. Immunoblots show the result of T3S assays in which proteins in culture supernatants (S, secreted proteins) and in bacterial pellets (P, non-secreted proteins) from ~$5x10^8$ and ~$5x10^7$ bacteria, respectively, were loaded per lane. The known *C. trachomatis* T3S substrates CT082 (312,313) and CT694 (266) were used as positive controls, and the *C. trachomatis* ribosomal protein RplJ was used as a negative control. SycO is a strictly cytosolic *Yersinia* T3S chaperone (295,299) and its immunodetection ensured that the presence of HA-tagged proteins in the culture supernatants was not a result of bacterial lysis or contamination. (C) The percentage (%) of secretion of each protein by *Y. enterocolitica* ΔHOPEMT was calculated by densitometry, as the ratio between the amount of secreted and total protein. The threshold to decide whether a protein was secreted was set to 2% (dashed line), based on the % of secretion of RplJ-HA. Data are the mean ± SEM from at least 3 independent experiments.

4.4.5. CT053, CT105, CT142, CT143, CT161, CT338, and CT429 can be translocated into host cells by *Y. enterocolitica*

We next analyzed if the newly identified likely T3S substrates of *C. trachomatis* had the capacity of being translocated into host cells, by using *Y. enterocolitica* as a heterologous system. For this, *Y. enterocolitica* ΔHOPEMT harboring plasmids encoding C-terminal HA-tagged newly identified likely T3S substrates of *C. trachomatis* (CT053-HA, CT105-HA, CT142-HA, CT143-HA, CT144-HA, CT161-HA, CT338-HA, CT429-HA, CT656-HA, or CT849-HA), a positive control (CT694-HA) or a negative control (RplJ-HA), were used to infect human epithelial HeLa cells. We then used Triton X-100 fractionation of the infected cells followed by immunoblotting analysis of Triton-soluble and insoluble HeLa cell lysates to monitor protein translocation into host cells. As expected, we found CT694-HA in the Triton-soluble fraction, which showed that this protein was delivered into the cytoplasm of HeLa cells, and only detected RplJ-HA in the Triton-insoluble fraction (Figure 4.4), which confirmed that this protein remained within the bacteria (and that the fractionation procedure did not lyse the bacteria). Among the 10 likely T3S substrates of *C. trachomatis* under analysis, we could not detect CT656-HA or CT849-HA in both the Triton-soluble and Triton-insoluble fractions. It is possible that in the experimental conditions used in this study CT656-HA or CT849-HA are translocated in minute and undetectable amounts and/or that they are degraded either after translocation or within the bacteria. Regardless of the exact scenario, these results did not enable us to conclude about the capacity of CT656-HA and CT849-HA of being translocated into host cells. However, we could consistently detect CT053-HA, CT105-HA, CT142-HA, CT143-HA, CT161-HA, CT338-HA and CT429-HA in the Triton-soluble fraction (Figure 4.4), indicating that these proteins were injected into the cytoplasm of HeLa cells by *Y. enterocolitica*. We could also occasionally detect small amounts of CT144-HA in the Triton-soluble fraction (barely visible in Figure 4.4).

In summary, these experiments showed that CT053-HA, CT105-HA, CT142-HA, CT143-HA, CT161-HA, CT338-HA and CT429-HA have the capacity of being translocated into infected host cells further suggesting that the endogenous *C. trachomatis* proteins could be effectors. The results do not preclude that CT144, CT656 or CT849 could be effectors, but the evidence is not as strong as for the other 7 proteins.

4.4.6. Expression of genes encoding newly identified likely T3S substrates during development of *C. trachomatis* LGV strain L2/434

To test if the newly identified likely T3S substrates, and possible effectors, of *C. trachomatis* (CT053, CT105, CT142, CT143, CT144, CT161, CT338, CT429, CT656, and CT849) were expressed during infection, and to gain insights of when they could be acting during the developmental cycle, we analyzed by RT-qPCR the mRNA levels of their encoding genes during the developmental cycle of strain L2/434, at 2, 6, 12, 20, 30 and 42 h post-infection.

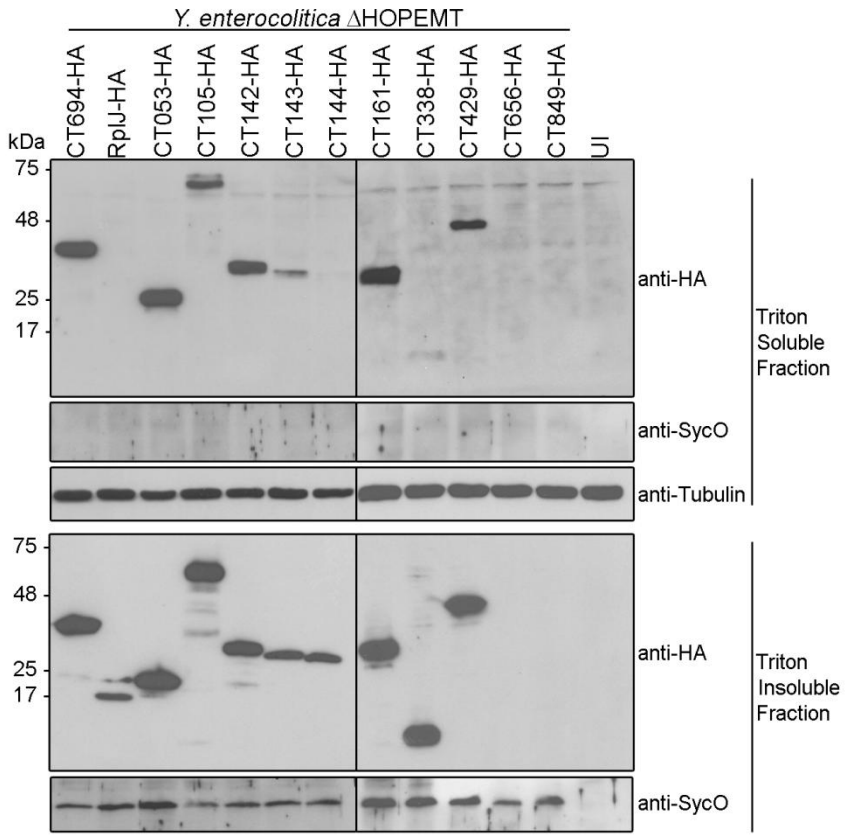


Figure 4.4. Translocation of *C. trachomatis* proteins into the cytoplasm of HeLa cells by *Y. enterocolitica*. HeLa cells were left uninfected (UI) or infected with *Y. enterocolitica* ΔHOPEMT strains expressing the indicated HA-tagged proteins. After 3 h of infection, extracellular bacteria were killed by the addition of gentamicin and the infected cells were incubated for additional 2 h. The infected cells were then fractionated into Triton-soluble and Triton-insoluble cell lysates that were subsequently analyzed by immunoblotting using anti-HA, anti-SycO and anti-tubulin antibodies, as indicated. Presence of HA-tagged proteins in the Triton-soluble cell lysates is indicative of translocation into the cytoplasm of HeLa cells. SycO is a strictly cytosolic *Yersinia* T3S chaperone (295,299) and its immunodetection ensured that the presence of HA-tagged proteins in the Triton-soluble cell lysates was not a result of bacterial lysis during the fractionation. Additionally, the incapacity to detect HA-tagged RplJ (a *C. trachomatis* ribosomal protein) in the Triton-soluble cell lysates further indicated that this fraction did not contain bacteria or non-translocated bacterial proteins. Tubulin served as a loading control of the Triton-soluble cell lysates. The images shown are representative of three independent experiments.

While *ct053*, *ct105*, *ct142*, *ct143*, *ct144*, *ct338*, *ct429*, *ct656*, and *ct849* displayed significant mRNA levels in more than one of the time-points analyzed, *ct161* showed only vestigial levels of expression throughout the cycle (Figure 4.5). The mRNA levels of *ct105* and *ct338* were > 5-fold higher at 2-6 h post-infection than in any other of the time-points analyzed (Figure 4.5), suggesting that the encoded proteins should function at early-cycle. The mRNA levels of *ct053* and *ct429* were higher between 6 and 20 h post-infection (Figure 4.5), suggesting that the encoded proteins might act from early to mid

cycle. The mRNA levels of *ct142*, *ct143*, *ct144* and *ct849* were higher at the later time points analyzed (30-42 h post-infection). However, while *ct142*, *ct143*, and *ct144* were expressed at similar levels at 30 and 42 h post-infection, *ct849* showed a distinct peak of expression at 30 h post-infection (Figure 4.5). Therefore, CT142, CT143, CT144 could function either at late or early cycle, and CT849 might probably acts at late cycle. Finally, the mRNA levels of *ct656* were constant at all time-points analyzed (Figure 4.5), suggesting that CT656 could function throughout the cycle. Regarding *ct161*, when comparing the higher mRNA levels detected for each of the genes analyzed, those of *ct161* were > 6-fold lower than those of any of the other genes tested (Figure 4.5). Therefore, in the experimental conditions used, CT161 may not be expressed by strain L2/434. In summary, the RT-qPCR experiments supported that CT053, CT105, CT142, CT143, CT338, and CT429, and also CT144, CT656, or CT849, could be *C. trachomatis* T3S effectors, possibly acting at different times of the developmental cycle.

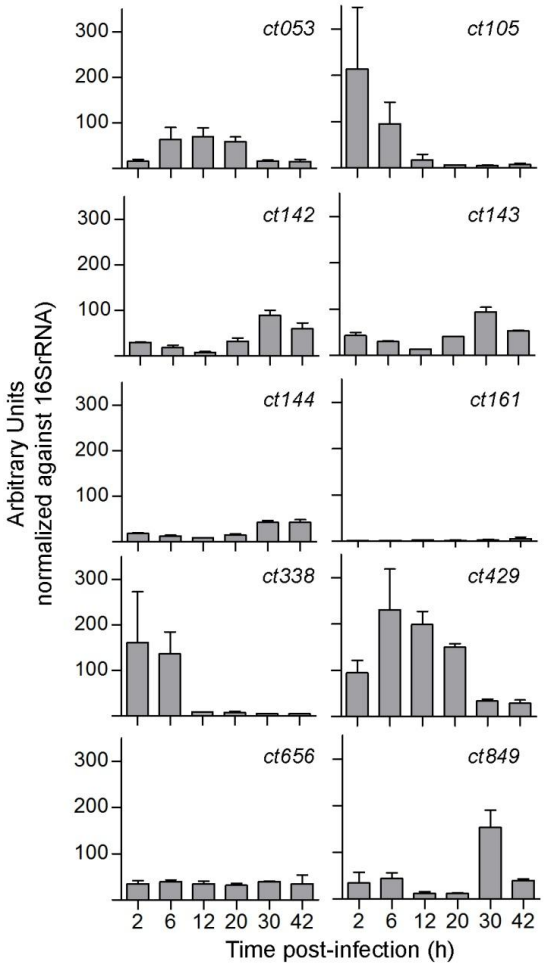


Figure 4.5. mRNA levels of newly identified putative effectors during the developmental cycle of *C. trachomatis* prototype strain L2/434. The mRNA levels of *ct053*, *ct105*, *ct142*, *ct143*, *ct144*, *ct161*, *ct338*, *ct429*, *ct656*, and *ct849* were analyzed by RT-qPCR during the developmental cycle of *C. trachomatis* strain L2/434, at the indicated time-points. The expression values (mean ± SEM) resulted from raw RT-qPCR data (10^5) of each gene normalized to that of the 16s rRNA gene and are from three independent experiments.

4.4.7. Comparison of the expression profiles of genes encoding newly identified likely T3S effectors during the development cycle of *C. trachomatis* strains displaying distinct tissue tropism

In order to evaluate whether changes in the gene expression of the newly identified likely T3S substrates, and possible effectors correlate with the tissue tropism of the strains, we scaled-up the gene expression analyses to additional strains representing the *C. trachomatis* ocular (strain C/TW-3) and epithelial-genital (strain E/Bour) disease-groups. This assay was performed for the genes *ct053*, *ct105*, *ct142*, *ct143*, *ct144* and *ct161*, given the observation that they encode proteins potentially translocated. Contrarily to *ct053*, where no considerable differences were observed across strains, the mRNA levels of the genes *ct142, ct143 and ct144* at mid- to late cycle were lower for C/TW-3 than for the remaining strains. Notably, the expression profiles of those three contiguous genes were highly similar across the strains, indicating that they are likely part of the same transcriptional unit. Also, *ct105* revealed differential expression features that may correlate with the tissue tropism of the strains (Figure 4.6). In fact, it displayed remarkably lower gene expression levels for the ocular and epithelial-genital strains than for the LGV isolate. Regarding *ct161*, although it was more expressed for the ocular and epithelial-genital strains, this result should be eyed with cautious considering the extremely low expression levels, and the fact that *ct161* is a pseudogene for C/TW-3 (Chapter V).

4.4.8. Search for the genetic basis underlying the potential differential disease-group expression of *ct105*

Considering the results obtained for *ct105*, we further analyzed the genetic variability among *C. trachomatis* strains within its potential promoter region. The putative transcriptional start site (TSS) of *ct105* was previously described by Albrecht et al. (102). Remarkably, we detected a 74-bp insertion 23 bp upstream from the predicted TSS of *ct105* in the L2/434 strain, when comparing with sequences from the strains C/TW-3 and E/Bour (Figure 4.7). Subsequent BLAST analyses revealed that this insertion is LGV-specific, as it perfectly discriminates all LGV from non-LGV strains (Figure 4.7). If the putative TSS described by Albrecht et al (102) is conserved among *C. trachomatis* strains, this LGV-specific insertion will imply a distinct -35 region of *ct105* for LGV strains. Although we did not evaluate the *ct105* expression for more LGV and non-LGV strains, we believe that *ct105* may display a LGV-specific expression pattern, potentially justified by the detected 74-bp LGV-specific insertion upstream from the TSS.

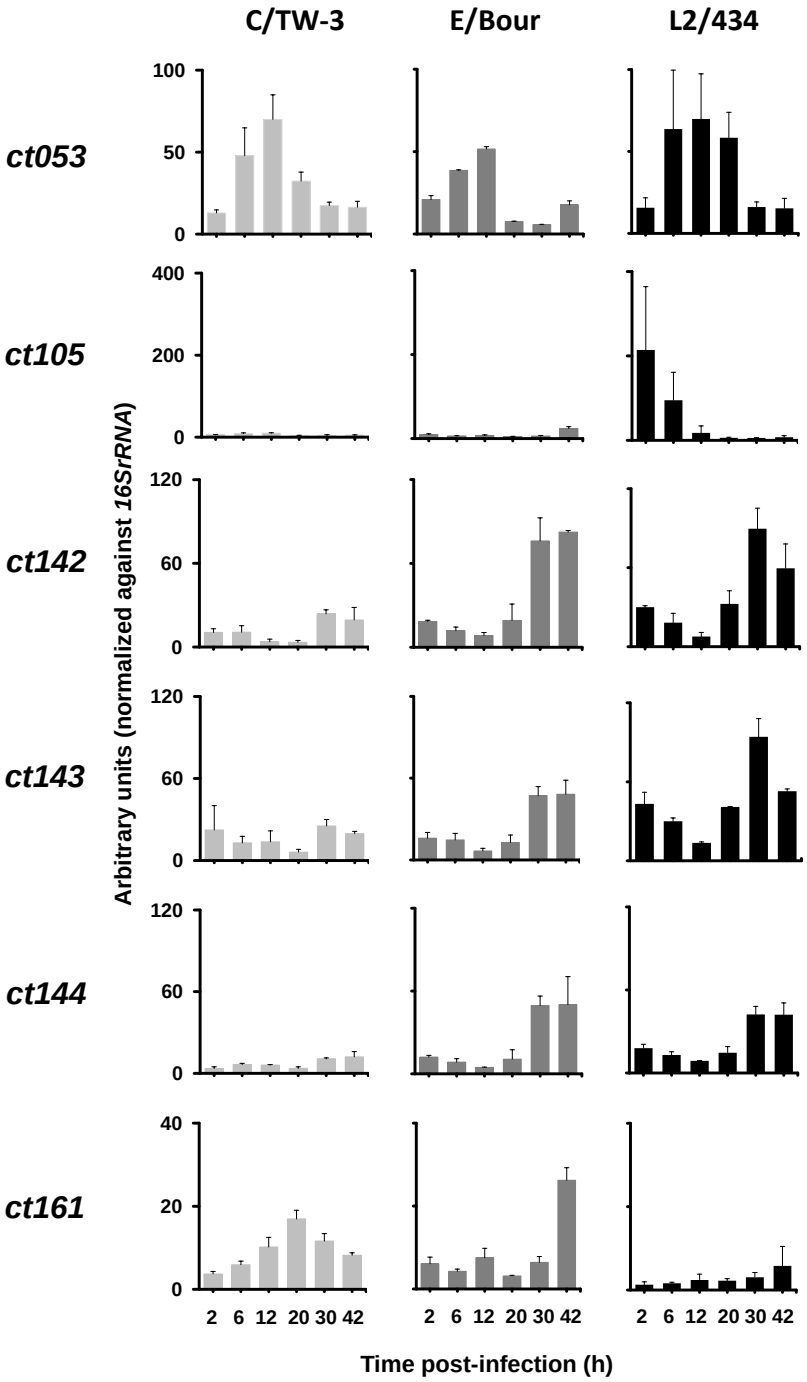


Figure 4.6. Comparison of the expression profiles of newly identified genes encoding putative effectors during the developmental cycle of *C. trachomatis* ocular, epithelial-genital and LGV strains. The mRNA levels of *ct053*, *ct105*, *ct142, ct143, ct144* and *ct161* were analyzed by RT-qPCR during the developmental cycle of the strains representing the three *C. trachomatis* disease groups: ocular (C/TW-3), epithelial-genital (E/Bour) and LGV (L2/434), at the indicated time-points. The expression values (mean ± SEM) resulted from raw RT-qPCR data (10^5) of each gene normalized to that of the 16s rRNA gene. Scales were adjusted for better visualization purposes.

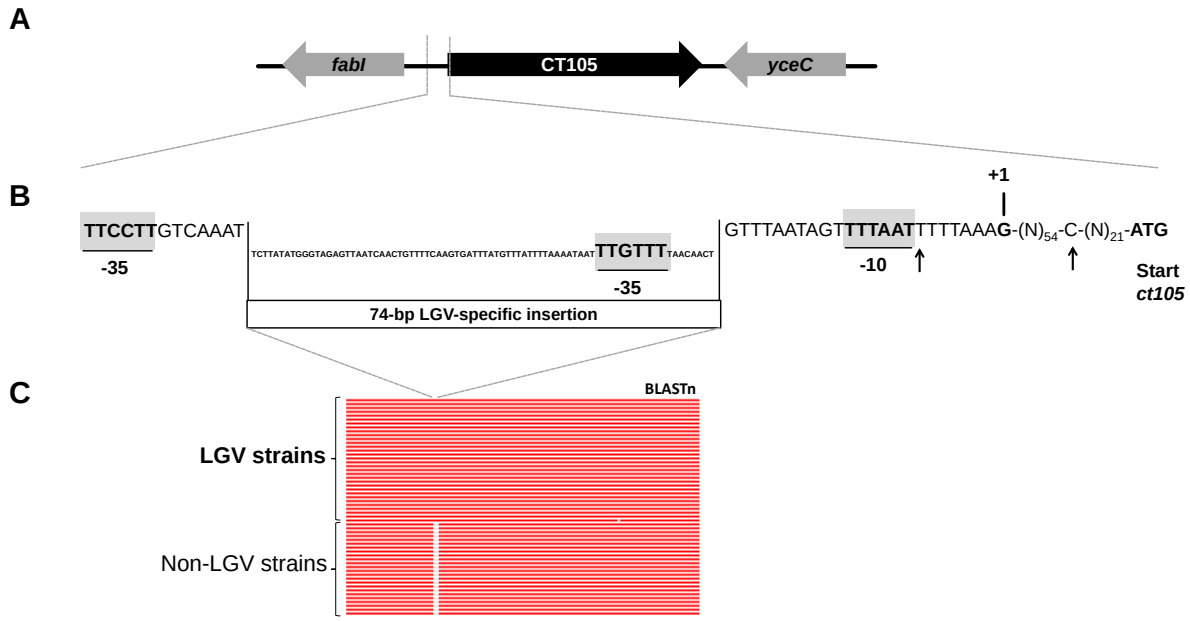


Figure 4.7. Identification of LGV-specific genetic features in the putative promoter region of *ct105*. A. Genetic organization of *ct105*. B. Schematic view of the nucleotide sequences of putative promoter region of *ct105*, highlighting the 74-bp LGV-specific insertion. The predicted transcription start site (102) is labeled by +1, and the predicted -10 hexamer and the potentially intra-species variable -35 hexamer are highlighted in grey. The analysis allowed the identification of additional LGV-specific nucleotides (indicated with an arrow). C. BLASTn analyses of the genomic region enrolling the *ct105* and flanking genes showing that the 74-bp insertion perfectly discriminates all LGV strains from non-LGV strains with sequences available at the GenBank.

4.5. Discussion

Earlier studies using heterologous systems have led to the identification of several *bona-fide* or putative *C. trachomatis* T3S effectors (132,140,266,290,291,309,310,312,313). While these and other analyses covered a significant portion of all *C. trachomatis* proteins, we hypothesized that there could be previously unidentified T3S substrates. By combining basic bioinformatics searches, exhaustive T3S assays, translocation assays, and analyses of chlamydial gene expression in infected cells, we revealed 10 *C. trachomatis* proteins (CT053, CT105, CT142, CT143, CT144, CT161, CT338, CT429, CT656, and CT849) as likely T3S substrates and possible effectors. In particular, CT053, CT105, CT142, CT143, CT338, and CT429 were type III secreted by *Y. enterocolitica*, could be translocated into host cells, and their encoding genes were clearly expressed in *C. trachomatis* strain L2/434. Therefore, these 6 proteins have a high likelihood of being effectors. However, additional future studies are required to show that all of these 10 proteins are indeed translocated by *C. trachomatis* into host cells and to show that they are *bona-fide* effectors, i.e., that they interfere with host cell processes.

Among the likely T3S effectors of *C. trachomatis* that we identified, CT105 and CT142 have been previously singled out as possible modulators of host cell functions, based on the phenotypic

consequences of their ectopic expression in yeast *S. cerevisiae* (156). Regarding *ct105*, it is worth noting that we have previously shown that it is a pseudogene for ocular strains (301) and that there is an overrepresentation of non-synonymous mutations when comparing sequences between genital and LGV strains (301), indicating that this gene is likely evolving following a niche-specific pattern. Additionally, we have now detected a LGV-specific 74-bp insertion within its promoter region that may potentially justify its higher expression in the L2/434 strain. Thus, besides the hypothesis that we may be facing a scenario of a gene needed for tropism functions other than those involving the ocular conjunctiva, we found genetic dissimilarities that may influence its expression in a disease group-specific manner. In addition, the genes encoding CT142, CT143, and CT144 have been shown to be markedly transcriptionally regulated by a protein (Pgp4) encoded by the *Chlamydia* virulence plasmid (207). This plasmid is present in almost all *C. trachomatis* clinical isolates (336), and studies in animal models of infection showed that it is a virulence factor *in vivo* (337,338). Additional studies are needed to understand if the putative effector function of CT142, CT143, and CT144 can partially explain the virulence role of the chlamydial plasmid. Our results of gene expression suggest that they are coordinately expressed regardless the tissue tropism of the strain and that they potentially belong to the same operon. The predicted amino acid sequence of CT849 reveals a domain of unknown function (DUF720) that can only be found in *Chlamydia* proteins. In *C. trachomatis*, besides CT849, a DUF720 domain is found in CT847, a T3S effector that interacts with human Grap2 cyclin D-interacting protein (GCIP) (290), and in CT848, which has been indicated as a T3S substrate using *S. flexneri* as a heterologous system (132). Therefore, this further supports a possible role of CT849 as an effector. In contrast with CT105, CT142, CT143, CT144 or CT849, no significant information is available or could be retrieved about CT053, CT338, CT429, or CT656.

CT161 is a possible T3S substrate and effector, but we could not detect significant levels of *ct161* mRNA during the developmental cycle of strain L2/434, and the levels of expression in the ocular and genital strains, although higher, are still very low when comparing with the other studied genes. Curiously, the *ct161* gene is localized within the “plasticity zone”, a chromosomal region of rare high genetic diversity among *C. trachomatis* strains, and CT161 has been shown by yeast two-hybrid to bind CT274 (a possible chlamydial T3S chaperone) (339). Another feature of this protein is that part of its amino acid sequence (residues 40-224, out of 246) shows 28% of identity to a region of Lda2/CT163 (residues 167-361, out of 548), a known *C. trachomatis* translocated protein (318).

Among the proteins for which we found a secretion signal but could not demonstrate their T3S as full-length proteins, we highlight CT153 and GrgA/CT504. Regarding CT153, this protein possesses a membrane attack complex/perforin (MACPF) domain (340), and there is previous evidence that it may be translocated by *C. trachomatis* (341), which is consistent with our data. The *ct504* gene has been recently shown to encode a transcriptional activator, GrgA (327). Therefore, T3S of $CT504_{20}$-TEM-1 could be false positive. However, if GrgA is a T3S substrate, as our data suggests, it could have a function within the host cell or, more likely and similarly to what has been described in the T3SSs of *Yersinia* (342) or *Shigella* (343,344), it could be discarded by secretion once its intra-bacterial regulatory activity needs to be shut down.

We found T3S signals in 56% proteins analyzed (26 out of 46, including controls). This high percentage of proteins showing a T3S signal suggests that some should be false positives. It is conceivable that within a single bacterium non-secreted proteins possess T3S signals but are not targeted to the T3SS machinery because they also carry signals (e.g. DNA-, membrane-, or protein-binding) that preferentially direct them to other location within the bacterial cell. To help differentiating between true or false positives among chlamydial proteins carrying a T3S signal we analyzed their secretion as full-length proteins. This is because, as explained above in the Results section, not all proteins have folding characteristics compatible with T3S (331-334). However, we cannot exclude that some of the *C. trachomatis* full-length proteins that were not type III secreted by *Yersinia* have a T3S chaperone that maintains them in a secretion-competent state (335) and enables their secretion during infection by *C. trachomatis*. Intriguingly, CT082 or CT694 have dedicated T3S chaperones, CT584 and Slc1, respectively (312), and, in agreement with what we previously observed (312), they were both secreted as full-length proteins in the absence of the chaperones. Considering that T3S chaperones have various functions (345,346), the chaperone role of CT584 or Slc1 should be different from maintaining their substrates in a secretion-competent state.

Eleven of the *Chlamydia* proteins that we analyzed have been previously studied for T3S using *S. flexneri* has a heterologous system (132). In the majority of the cases the outcome of the experiments was identical; however, differently from what was shown in *Shigella*, we detected a T3S signal in the N-terminal of CT429 (which was also secreted as a full-length protein), GrgA/CT504, and CT779 and we did not detect a T3S signal in CT577. Evidence for a T3S signal in only one of the heterologous systems may suggest a false positive. There is a myriad of possible explanations for these discrepancies, when considering that different heterologous systems (*Shigella* and *Yersinia*) and reporter proteins (Cya and TEM-1) were used, and that the N-terminal regions in the hybrid proteins consisted in different lengths of amino acids and were in some cases from different *Chlamydia* species.

We compared the data from our T3S assays (including the controls, CT082, CT694, and RplJ) with predictions of T3S substrates by *in silico* methods [Effective T3S (314), SIEVE (315), Modlab (316), and T3_MM (328)] using resources available in the Web (Effective T3S, Modlab and T3_MM) and Table 3 in reference (315) (SIEVE), as detailed in the Supplemental Table S4.3.. When considering the analysis of T3S signals in TEM-1 hybrids, the vast majority of proteins (60%; 12 out of 20) in which we did not find a T3S signal were also predicted not to be secreted by each of the *in silico* methods. In contrast, the vast majority of proteins (58%; 15 out of 26) in which we detected a T3S signal were also predicted to be secreted by at least one of the *in silico* methods. The correlation between our experimental data and the *in silico* predictions was more striking when considering the T3S of full-length proteins. Among the 16 full-length proteins for which we could not find definitive evidence of T3S, 10 (i.e., 62.5%) were also predicted not to be secreted by each of the *in silico* methods, but among the 11 proteins that we showed or confirmed to be T3S substrates, 10 (i.e., 83%) were also predicted to be secreted by at least one of the *in silico* methods. Overall, this indicates some correlation between our experimental data and the *in silico* methods that predict T3S substrates. However, for many proteins, each of these *in silico* methods generates different predictions

(Supplemental Table S4.3). It is possible that the quantitative data on T3S such as the one we generated in this and in a previous study (324), can be used to normalize and improve the predictive value of such methods.

4.6. Conclusions

We found 10 *C. trachomatis* proteins (CT053, CT105, CT142, CT143, CT144, CT161, CT338, CT429, CT656, and CT849) with a high likelihood of being T3S substrates, and therefore possible effectors delivered by the bacteria into host cells. For 6 of these proteins (CT053, CT105, CT142, CT143, CT338, and CT429), the hypothesis that they could be effectors was supported by their capacity of being translocated into host cells and by the expression of their encoding genes by *C. trachomatis*. Additionally, gene expression analyses pointed out that they possibly act at different times of the developmental cycle, and some of them may contribute to the distinct tropism of *C. trachomatis* strains. The identification of all *C. trachomatis* effectors is a crucial step towards a comprehensive understanding of the mechanisms by which this pathogen subverts host cells. The recently developed methods for genetic manipulation of *Chlamydia* indicate that it should be possible to ectopically express candidate effectors in *C. trachomatis* (206,212), which would facilitate the analysis of their translocation into host cells. Our work highlights *C. trachomatis* proteins that should be prioritized in such studies, thus aiding the future identification of chlamydial effectors. Furthermore, the quantitative analysis of T3S of TEM-1 hybrid proteins that we carried out could help to further develop the *in silico* methods for identification of T3S substrates (314-316,328).

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through grants ERA-PTG/0005/2010 (in the frame of ERA-NET PathoGenoMics) to LJM, ERA-PTG/0004/2010 (in the frame of ERA-NET PathoGenoMics) to JPG, and PEst-OE/EQB/LA0004/2011; by the European Commission through a Marie Curie European Re-integration Grant (PERG03-GA-2008-230954) to LJM; and by a European Society for Clinical Microbiology and Infectious Diseases (ESCMID) research grant to LJM. MdC, FA, and VB held PhD fellowships SFRH/BD/62728/2009, SFRH/BD/73545/2010, and SFRH/BD/68527/2010, respectively, from FCT.

Competing interests

All authors read and approved the final manuscript, declared that they have no competing interests.

Complete genome sequence of *Chlamydia trachomatis* ocular serovar C strain TW-3

The contents of this chapter were adapted from the following publication through the inclusion of additional data concerning genetic features revealed or confirmed upon fully sequencing and annotation of the genome of a C. trachomatis serovar C strain.

Borges V., Pinheiro M., Vieira L., Sampaio D.A., Nunes A., Borrego M.J. Gomes J.P.
Complete genome sequence of *Chlamydia trachomatis* ocular serovar C strain TW-3.
2014, Genome Announcements, 2:e01204-13.
(http://www.ncbi.nlm.nih.gov/pubmed/24459269)

Personal Contribution

VB contributed to the design of the study, performed most of the experimental work and bioinformatics analyses, interpreted data and wrote the manuscript.

5. Complete genome sequence of *Chlamydia trachomatis* ocular serovar C strain TW-3

Abstract

Chlamydia trachomatis is the etiological agent of trachoma, the leading infectious cause of blindness worldwide. We report here the first complete and annotated genome of a *C. trachomatis* trachoma-causing serovar C strain (strain TW-3). The chromosome and plasmid are 1,043,554 bp and 7,501 bp in length, respectively.

The obligate intracellular bacterium *Chlamydia trachomatis* is the major cause of bacterial sexually transmitted infections worldwide, and is also responsible for trachoma, the leading infectious cause of blindness. Trachoma results from a conjunctival chronic inflammatory state leading to the formation of irreversible corneal opacities and blindness (41,54). It is one of 17 neglected tropical diseases and is targeted for elimination by 2020 by the World Health Organization (WHO) through the implementation of the SAFE strategy: lid surgery (S), antibiotics to treat the infection (A), facial cleanliness (F); and environmental changes (E) (55). Trachoma affects more than 40 million people (being responsible for the visual impairment of about 2.2 million people) and it is endemic in more than 50 countries, predominantly in sub-Saharan Africa, the Middle East, and Asia (54,347). This epidemiological pattern is essentially associated with *C. trachomatis* ocular serovars A and B (53) (from the existent 15 major serovars) whereas serovar C seems relatively common in indigenous Australian communities (52,348). Of note, serovar C has been also associated with *Chlamydia*-related arthritis (76). There are already five and two annotated genomes from serovars A and B (37,165,166), respectively. We report here the first complete and annotated sequence of a trachoma-causing *C. trachomatis* serovar C strain.

The C/TW-3 strain was isolated in Taiwan in 1959 from the human conjunctiva (349). We obtained this strain from the American Type Culture Collection (ATCC VR-1477), and propagated it in HeLa229 cells monolayers before proceeding with bacterial purification using discontinuous density gradients (350). The whole-genome sequence was determined by using a paired-end strategy (2 x 250 bp) with the platform Illumina MiSeq. Reads were mapped to *C. trachomatis* chromosome and plasmid sequences (37,165,166) by using both Bowtie 2 (version 2.1.0; [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml]) (351) and BWA software (version 0.7.5a; [http://bio-bwa.sourceforge.net/]) (352). Globally, 5,233,958 reads (with a mean quality score above 30 for > 95% of reads' bases) were mapped, which yielded a mean coverage of 1117-fold and 6717-fold for the chromosome and plasmid, respectively. SNPs and indels were identified using SAMtools followed by variant calling using BCFtools [http://samtools.sourceforge.net/] (353), and were carefully inspected through the Integrative Genomics Viewer (version 2.3.12; [http://www.broadinstitute.org/igv/]) (354). Both the typing gene (*ompA*) and problematic regions (ex. *tarp*) were confirmed by PCR followed by Sanger sequencing. The sequence was annotated by NCBI Prokaryotic Genomes Annotation Pipeline 2.3.

The genome sequence of C/TW-3 revealed a chromosome of 1,043,554 bp in length, with a G+C content of 41.30 % and 922 predicted CDSs. Plasmid analysis revealed the existence of about six copies *per* chromosome (based on the ratio of mean coverage for plasmid/chromosome), which fits previous data (355), and was found to be 7,501 bp in length comprising eight CDSs (G+C content of 36.25%).

Firstly, we analysed the tryptophan synthase operon (*trpRBA*) and the cytotoxin locus, since these loci carry well-established genetic features that distinguish *C. trachomatis* strains with dissimilar tissue tropism. As expected for typical oculotropic isolates, we confirmed that the C/TW-3 strain has a predicted non-functional *trpRBA* operon, since the gene encoding the tryptophan synthase α subunit (CT171/*trpA*) is disrupted (51,180,285). Also, it only retains the UDP-glucose binding domain of the cytotoxin (CT166) (the functional glycosiltransferase domain is missing), which implies the lack of enzymatic activity and, consequent, predicted lack of cytotoxicity in HeLa cells (179,187). We further noticed other genes containing mutations (SNPs or insertion/deletions) that prematurely truncate the predicted proteins (pseudogenes), namely the genes CT036, CT105, CT157, CT161, CT163, CT167, CT172, CT192, CT228, CT374/*aacX*, CT397/*vacB*, CT442/*crpA*, CT459/*pfrB*, CT593/*sdhC*, CT777/*bioF* and CT867/*ChlaDub2* (nomenclature according to the genome of the D/UW3). Unsurprisingly, five of these pseudogenes (CT157, CT161, CT163, CT167, CT172) [some of them previously annotated as pseudogenes in other *C. trachomatis* genomes (37,46,165,166)], are located in the so-called "plasticity zone". This is a ~20-25 kb region of the *C. trachomatis* genomes (locus ~CT152-CT176) known to carry not only a high level of genetic diversity within *Chlamydiaceae*, but also important genetic differences within the *C. trachomatis* species, including those occurring in the *trpRBA* and the cytotoxin loci (167,189). We also highlight CT192 and CT228, as they were previously described as pseudogenes in some serovar A and/or B strains (165,166), and both potentially encode putative inclusion membrane proteins (129,156,324). Although there are some issues regarding the annotation of the gene CT192 in the *C. trachomatis* genomes (324) and the genome of the serovar A prototype strain (A/Har13) encodes non-disrupted CT192 and CT228 genes (165), we hypothesized that these genes are being specifically targeted by relaxed selection in ocular strains. In fact, besides the well-known case of the *trpRBA* operon, this phenomenon (i.e. the existence of potential "ocular-specific" pseudogenes) was previously suggested to also target the genes CT058 (324), CT105 (301) and CT374/*aacX* (356) (the two later are also pseudogenes in the C/TW-3 strain). Finally, in order to integrate the whole genetic backbone of the *C. trachomatis* strain C/TW-3 in the frame of the known species phylogeny and diversity, we performed a phylogenetic reconstruction using additional 52 available chromosome sequences (37) from strains representing the three disease-groups of *C. trachomatis*. We observed the segregation of the C/TW-3 strain together with remaining ocular strains (Figure 5.1). Besides its inclusion in the "ocular branch", it differs more than 1100 SNPs from any ocular strain. These suggest that, contrarly to what is expected to the genital group, a more comprehensive knowlegde on the genetic variability within this disease-group will implicate the full-genome availability of several other ocular strains.

In conclusion, the availability of a complete and annotated genome sequence of an additional trachoma-causing serovar may contribute for the elucidation of the genetic basis underlying pathogenic differences among *C. trachomatis* ocular strains (76,164,357).

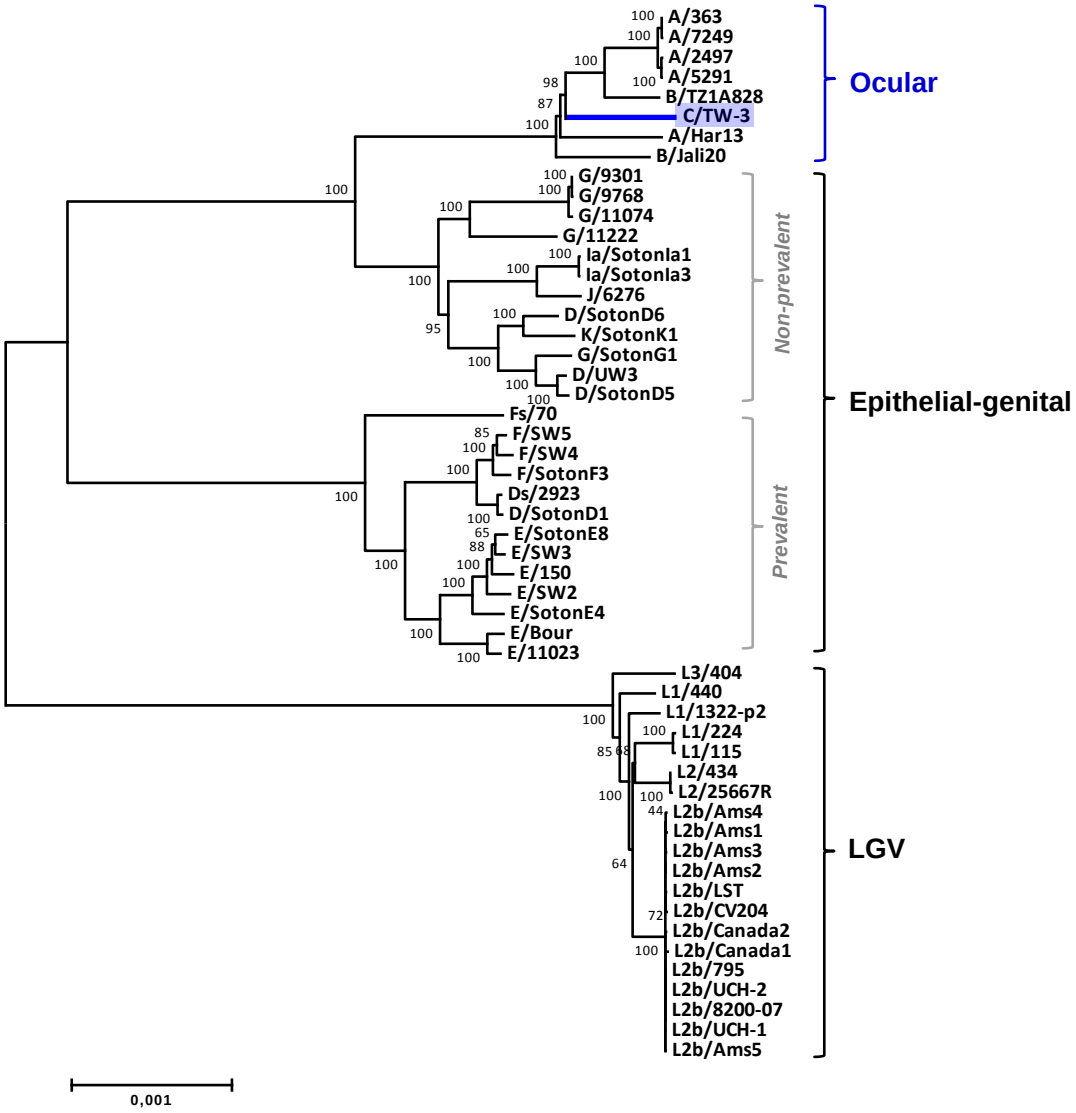


Figure 5.1. Integration of the genetic backbone of the sequenced and annotated genome of the *C. trachomatis* serovar C strain TW-3 in the species phylogeny and diversity. The "species tree" enrolls the complete chromosome sequences of 52 previously analyzed *C. trachomatis* strains (37) and the strain C/TW-3 (highlighted in blue). The tree topology shows the segregation of strains from ocular serovars (A-C), LGV serovars (L1-L3), and also, the existence of two distinct clades for the strains from the epithelial-genital serovars (D-K): one clade enrolling the most clinically prevalent serovars (mostly E and F strains) and one clade including the non-prevalent epithelial-genital serovars (37). The genomes were aligned using the progressive algorithm of Mauve software (version 2.3.1) (232), and a core-alignment (enrolling ~97% of the chromosome) was extracted by keeping regions where all chromosome sequences aligned over at least 500 bp. Phylogenetic estimations were carried out through MEGA5 by using the neighbor-joining method with bootstrapping (1000 replicates) (values are shown next to the branch nodes) (234), where the Kimura two-parameter method (235) was employed to estimate the evolutionary distances.

Nucleotide sequence accession numbers.

The complete genome sequence of *C. trachomatis* serovar C (strain TW-3) has been assigned GenBank accession numbers CP006945 (chromosome) and CP006946 (plasmid).

Acknowledgements

This work was supported by the grant ERA-PTG/0004/2010 from Fundação para a Ciência e a Tecnologia (FCT) in the frame of ERA-NET PathoGenoMics (to JPG). VB and AN were recipients of a Ph.D. fellowship (SFRH/BD/68527/2010) and a post-doctoral fellowship (SFRH/BPD/75295/2010), respectively, from FCT.

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

Manuscript with minor changes published in

2015, Infection, Genetics and Evolution, 32:74-88.
Borges V., Gomes J.P.
Deep comparative genomics among *Chlamydia trachomatis* lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation
(http://www.ncbi.nlm.nih.gov/pubmed/25745888)

These results were presented as a poster in the Seventh Biennial Meeting of the *Chlamydia* Basic Research Society, New Orleans (Louisiana, USA), March 29 - April 1, 2015.

Personal contribution

VB contributed to the design of the study, performed all the experimental work and bioinformatics analyses, interpreted data and wrote the manuscript.

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

6.1. Abstract

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

6.2. Introduction

Lymphogranuloma venereum (LGV) is a human sexually transmitted disease caused by strains from the serovars L1 to L3 of the obligate intracellular bacterium *Chlamydia trachomatis*. Upon genital or rectal contact, these strains have the ability to infect mononuclear phagocytes and spread into regional lymph nodes causing an inguinal syndrome characterized by genital ulcers and painful inguinal lymphadenopathy (the typical LGV bubonic form) (45,46). Therefore, LGV strains have been historically considered to represent a biovar (LGV biovar) distinct from the trachoma biovar that is comprised by strains from the ocular serovars (A to C) and genital serovars (D to K), which otherwise are normally associated with infections restricted to the epithelial cells of the conjunctival and genital mucosae, respectively. The ocular strains are the causative agent of trachoma, the world's leading

cause of preventable infectious blindness (6), whereas *C. trachomatis* strains from serovars D to K are the major cause of bacterial sexually transmitted diseases worldwide (4).

Although cases of typical bubonic LGV presentation have been endemic in parts of Africa, Asia, South America and the Caribbean but are rare in the Western World (5,45,49,358-361), an atypical LGV clinical presentation characterized by severe ulcerative proctitis (so called anorectal syndrome) has emerged in Europe and North America since 2003 (5,48,49,67,362,363). This anorectal syndrome, which mostly afflicts men who have sex with men (MSM), normally co-infected with HIV and other sexually transmitted infections (66,67), was found to be primarily caused by strains from the LGV serovar L2b. The genome sequencing of epidemic L2b isolates causing proctitis revealed that these isolates are nearly clonal (37), and carry no evident genomic particularities (e.g., gene gain/loss) that could unequivocally be associated with the atypical symptoms (46). Nonetheless, although sporadically, L2b infections may progress to inguinal lymphadenopathy or persistent genital ulcers (5,67-69), and L2b-associated cases of proctitis and bubonic LGV in a woman have been described (66,70,71), suggesting that L2b strains may carry specific genetic features that enable them to exhibit wider tropism and transmission skills than strains from the remainder LGV serovars. These data, together with the existence of LGV strains with rather unusual recombinant profiles (175), clearly suggest that the diversity among LGV strains is higher than expected (37), and thus, point out that more sampling and deeper molecular characterization of LGV strains are needed to better understand the pathodiversity within the LGV biovar. For instance, comparative genomic analysis of *C. trachomatis* ocular strains was important to find subtle genomic variations likely underlying virulence dissimilarities (164), and we have previously identified some genes targeted by positive selection events likely driving phenotypic diversity among LGV strains (301). Besides host-related factors (genetics and/or immunological) [reviewed in Abdelsamed et al. (146) and Asner et al. (147)], the differential disease severity and tissue tropism among LGV strains is likely a function of the genetic backbone of the strain, where few strain- or serovar-specific mutations may be determinant factors for particular phenotypic signatures. In fact, although LGV strains are considerably more related than those of the trachoma biovar and discreet recombination within the LGV biovar has occurred, hundreds of mutations are known to be fixed since the LGV ancestral lineage diverged from the remaining *C. trachomatis* strains (37,173,301). In addition, the evolutionary pathway underlying the L2b clonal expansion and dispersion in MSM is still to be deciphered. Considering this scenario, we intended to deeply characterize the genetic variability among LGV strains and to get insight into the mutational events that occurred throughout the pathoadaptation of the epidemic L2b lineage.

6.3. Materials and Methods

6.3.1. *C. trachomatis* strains, cell culture and bacterial DNA purification

Two serovar L2b *C. trachomatis* clinical strains [*ompA* genotype confirmed as previously described (277)] were isolated from positive diagnostic specimens collected from MSM suspected of having LGV proctitis that attended to the major Portuguese sexually transmitted disease clinic (Lapa

Health Centre, Lisbon). The L2b/CS19/08 strain was isolated from a anoarectal swab of a 29-years old MSM with proctitis, syphilis and HIV(+), whereas the L2b/CS784/08 strain was collected from an anoarectal swab of a 32-years old MSM having anal bloody discharge and being both HIV and *Treponema pallidum* negative. Each clinical specimen in transport medium (2 sucrose phosphate buffer supplemented with gentamicin, vancomycin and nistatin) was inoculated (0.2 ml per well) in HeLa229 confluent monolayers (cultured on 24-well plates in MEM containing 10% fetal bovine serum, 5 mM L-glutamic acid, 10 µg/ml gentamicin and 0.5 µg/ml fungizone at 37 ºC, 5% CO_2) by centrifuging at 3500 rpm for 1h at 34 ºC. The cultures were subsequently incubated for 1h at 37 ºC, 5% CO_2, and the cell medium was replaced by fresh medium supplemented with vitamins (1x), non-essential aminoacids (1x), glucose (0.5%) and cycloheximide (0.5 µg/ml). Cultures were allowed to grow at 37 ºC, 5%CO_2 until about 48 h post-infection. The yield of infection was monitored by immunofluorescence microscopy after fixing cultures with methanol and staining with an anti–*C. trachomatis* lipopolysaccharide antibody (Pathfinder), according to manufacturer's instructions. The medium was then removed, the bacterial-infected cells were harvested using glass beads and disrupted through sonication (Vibra Cell, Bioblock Scientific), and further submitted to low-speed centrifugation (700 rpm for 7min). The bacterial-enriched supernatant was collected, homogenized and re-inoculated onto new HeLa229 confluent monolayers, as described above. Both chlamydial cultures were scale-up (6-7 passages) until obtaining eight T25 cm^2 flasks containing chlamydial-enriched HeLa229 monolayers, and further subjected to a discontinuous density gradient purification procedure, as previously described (350). DNA was finally extracted from the elementary bodies (EBs) fraction using the DNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions, and further quantified by applying the Quant-iT™ PicoGreen® dsDNA Assay (Invitrogen). The DNA integrity was assessed through agarose gel electrophoresis.

6.3.2. Whole-genome sequencing, assembly and alignment

High-quality DNA samples were used to prepare Illumina paired-end libraries according to the manufacturer's instructions (by using Nextera XT DNA sample preparation and index kits; Illumina Inc., San Diego, CA). DNA libraries were further loaded onto MiSeq reagent cartridge (MiSeq Reagent Kit v2; Illumina Inc., San Diego, CA, USA) and, subsequently, subjected to cluster generation and paired-end sequencing (2x250 bp) on a MiSeq platform (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. Illumina reads were mapped to chromosome and plasmid sequences from the *C. trachomatis* strain L2b/UCH-1 (GenBank accession numbers AM884177 and AM886279) (46) using both Bowtie2 (version 2.1.0 [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml]) (351) and the Burrows-Wheeler Aligner (BWA) software (version 0.7.5a [http://bio-bwa.sourceforge.net/]) (352). The reference-based approach seemed accurate as the epidemic L2b isolates are expectedly clonal and carry no gene gain/loss differences (37,46). So, the chance of finding new, unique sequences for these two new genomes that did not align to L2b/UCH-1 is virtually zero. The obtained mean depth coverage for the L2b/CS19/08 and L2b/CS784/08 strains were 68-fold and 26-fold (for the chromosome) and 586-fold and 226-fold (for

the plasmid), respectively; with no single regions displaying zero coverage. Based on the ratio plasmid/chromosome taken from the respective depth coverage, we were able to infer the plasmid copy number *per* chromosome (about nine for both strains). SAMtools/BCFtools (http://samtools.sourceforge.net/) (353) were applied to call Single Nucleotide Polymorphisms (SNPs) and indels, which were confirmed through visual inspection using the Integrative Genomics Viewer (version 2.3.12 [http://www.broadinstitute.org/igv]) (354). Chromosome and plasmid sequences were annotated by the NCBI Prokaryotic Genomes Annotation Pipeline 2.3, and further assigned under the GenBank accession numbers: CP009923 and CP009924 (for the L2b/CS19/08 strain); and CP009925 and CP009926 (for the L2b/CS784/08 strain); respectively. For both strains, the plasmid size (7500 bp) matches the one observed for the reference strain L2b/UCH-1 and chromosome size (1 038 864 bp) differs by solely one base pair.

6.3.3. Analysis of genetic diversity among LGV strains: mapping of all variant sites and indels

Comparative whole-genome analyses were performed to identify all genetic differences among *C. trachomatis* LGV strains. We analyzed a set of 22 strains (Supplemental Table S6.1), which included four L1, two L2 and one L3 strain, as well as fifteen serovar L2b strains from the recent L2b epidemics in Europe and North America. The two L2b strains (L2b/CS19/08 and L2b/CS784/08) isolated in Portugal are the first isolates (clinically associated with the ongoing LGV outbreak) from Western Europe subjected to whole-genome sequencing. Although a hybrid LGV *C. trachomatis* strain originated from recombination between serovar L2 and D lineages has been reported (175), we opted not to include it in the present study as its uniqueness and rather unusual profile of inter-biovar recombination would markedly bias the interpretation of the results. Chromosome and plasmid DNA sequences were aligned using the progressive algorithm of Mauve software (version 2.3.1) (232), and the list of all variant sites (i.e., sites in the genome that have a SNP in at least one strain) and indels among LGV strains were extracted. In order to improve the alignment quality and reduce misleading comparisons, we constructed several alignments involving sets of sequences from different strains (such as all LGV strains, all L2b strains, or the plasmid-bearing L2/434/Bu strain *versus* the naturally occurring plasmidless L2/25667R strain). Even so, all mutation events were further checked and carefully confirmed after alignment correction and visual inspection through MEGA5 software (http://www.megasoftware.net) (233). Mutation events in regions with unequivocal annotations bias (i.e., marked by strings of “N”) were not validated. DnaSP v5 software (364) was also applied to determine the SNP density across each alignment. Globally, we were able to identify all variant sites and indels differentiating LGV strains, and to categorize them according to their location [gene or intergenic region (IGR)] and discriminatory role as: *i)* serovar-specific mutation (i.e., distinguish a singular serovar from the remainder LGV serovars); *ii)* mutation involved in L2b clone radiation (i.e., mutation among L2b strains); or *iii)* mutation differentiating a plasmid-bearing strain (L2/434/Bu) from a same-serovar plasmidless strain (L2/25667R). Additionally, all variant sites located in coding regions were classified as nonsynonymous or synonymous through detailed analyses of each individual gene

alignment using Lasergene 9.0 tools (DNASTAR, Madison, WI) and MEGA5. SNPs occurring in regions annotated as pseudogenes in the L2/434/Bu strain annotation or with strikingly discordant annotations among LGV strains (Supplemental Table S6.2) were excluded from the nonsynonymous/synonymous classification. We also evaluated the possibility of the SNPs occurring in coding regions to be homoplasic relative to strains from the trachoma biovar. Phylogenetic estimations were conducted through MEGA5 by using the neighbor-joining method with bootstrapping (1000 replicates) (234), where the Kimura two-parameter method (235) was employed to estimate the evolutionary distances. MEGA5 was also applied to estimate the nonsynonymous/synonymous substitution rate ratio (*dN*/*dS*) among sequences from LGV strains, where *dN* corresponds to the number of nonsynonymous substitutions *per* nonsynonymous site and *dS* is defined as the number of synonymous substitutions *per* synonymous site. For specific figures or tables, we used a dual locus designation ("CT" or "CTL") based on the genome annotations of the *C. trachomatis* strains D/UW-3/CX (GenBank accession number NC_000117) and L2/434/Bu (GenBank accession number NC_010287) in order to allow the reader a better association with literature data.

6.4. Results and Discussion

6.4.1. Global genetic diversity among *C. trachomatis* LGV isolates

In a previous study, we showed that about 80% of all *C. trachomatis* genes phylogenetically distinguish LGV strains from the remainder strains from different disease groups, with almost one third segregating the LGV clade in an exclusive fashion (365). Here, we focused on the genetic variability among *C. trachomatis* LGV strains (Supplemental Table S6.1) in order to identify the genetic trends that may hypothetically underlie the dissimilarities in disease outcomes presented by these strains, providing a global view of the latest stages of the evolutionary pathway of the LGV biovar. Therefore, we performed comparative whole-genome analyses over 22 strains representing the *C. trachomatis* serovars that are known to be associated with the typical bubonic LGV presentation or with the anorectal syndrome with acute proctitis (current epidemic atypical LGV presentation). Globally, we found 453 genes and 144 IGRs displaying genetic differences (SNPs and/or indels) among the *C. trachomatis* LGV strains. By overlapping the set of genes involved in the phylogenetic separation of the LGV clade (previous study) with the ones detected to carry mutations among LGV strains (present study) (Figure 6.1 and Supplemental Table S6.2), we were able to find out that 79% of the latter also display a branch segregating all LGV strains from the trachoma biovar strains (44% if we considered solely the genes that exclusively segregate the LGV clade while the other strains are mixed). This group of genes was likely involved in the evolutionary segregation of the LGV clade and has been accumulating other mutations as a result of the never-ending evolutionary process of *C. trachomatis* adaptation to either mononuclear phagocytes or the rectal mucosa. On the other hand, about half of *C. trachomatis* genes responsible for the separation of the LGV clade are 100% conserved among these strains (Figure 6.1), suggesting some degree of purifying selection to maintain their function in the LGV biovar.

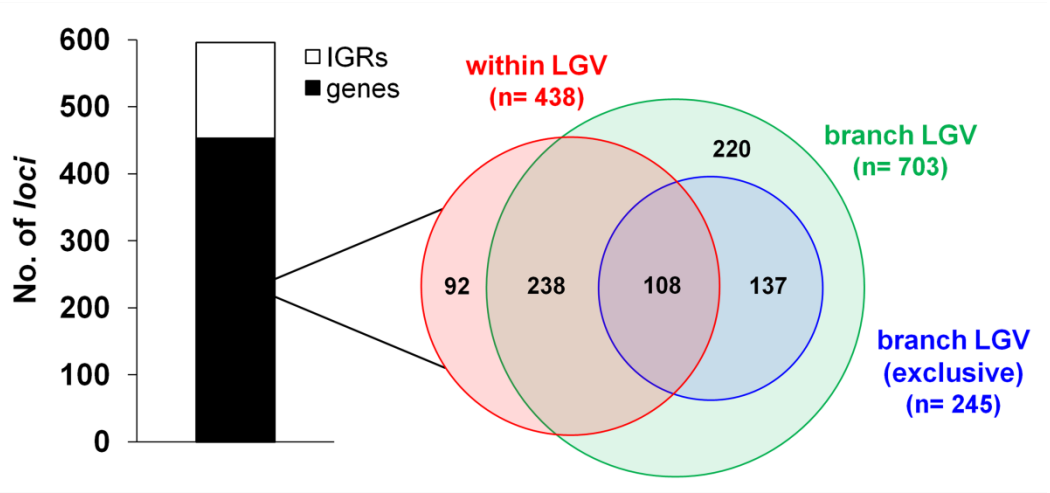


Figure 6.1. Genetic diversity among *C. trachomatis* LGV isolates. The bar graph shows the number of *C. trachomatis* genes (black) and IGRs (white) displaying genetic differences (SNPs and/or indels) among LGV strains. The Venn diagram represents the degree of overlapping between the set of genes involved in the phylogenetic separation of the LGV clade from the remainder strains (green circle) (365) and the group of genes identified to carry mutations among LGV strains in the present study (red circle). The amount of genes that exclusively segregate the LGV clade (i.e., cluster the LGV strains while the remainder strains, regardless of their phenotype, are mixed together) is also shown (blue circle). 346 out of the 438 genes (79%) varying among LGV strains present also the phylogenetic LGV clade (red circle ∩ green circle). From the 245 genes exclusively showing the phylogenetic LGV branch, 108 (44%) also display intra-LGV variability (red circle ∩ blue circle). The number of genes presented in the Venn diagram does not perfectly match the total number of genes detected in both studies, due to the exclusion criteria used in the two studies (implying, for instance, that the 438 and not 453 genes were used for this comparison). Circles were drawn proportionally to the number of genes each one represents.

Regarding the variable loci among LGV strains identified in the present study, we listed the total number of variant sites and indels, as well as their distribution according to their discriminatory role, as described above: serovar-specific mutations, mutations involved in L2b clone radiation, or mutations differentiating the plasmid-bearing strain (L2/434/Bu) and the same-serovar naturally occurring plasmidless strain (L2/25667R) (Supplemental Table S6.2). The variant sites occurring in coding regions were also classified as nonsynonymous or synonymous. The global statistics are summarized in Table 6.1. Some discrete data from the global trends presented in this section have previously been reported by Harris and colleagues (37), although they have not specifically focused on LGV group. Here, we present a deeper global analysis exclusively focused on these strains, which will be fundamental for the understanding of the following sections.

Table 6.1. Global trends of genetic variability among LGV strains.

	Total	**Total (w/o *ompA*)**	N [a]	**N [a] (w/o *ompA*)**	S [b]	**S [b] (w/o *ompA*)**	H [c, e]	**H [c, e] (w/o *ompA*)**	Indel [d, e]	**Indel [d, e] (w/o *ompA*)**
Variant sites among LGV strains	1497	**1279**	714	**607**	597	**486**	~390	**~180**	~100	**~100**
L1-specific	14	**6**	9	**2**	3	**2**	8	**0**	2	**1**
L2-specific	133	**133**	54	**54**	45	**45**	7	**7**	6	**6**
L2b-specific	98	**97**	54	**53**	28	**28**	5	**4**	11	**11**
L3-specific	312	**162**	142	**67**	139	**64**	164	**18**	14	**12**
Variant sites involved in L2b clone radiation	44	**44**	37	**37**	4	**4**	2	**2**	7	**7**
Variant sites differentiating L2/434 (plasmid-bearing) from L2/25667R (plasmidless)	8	**8**	6	**6**	0	**0**	0	**0**	1	**1**

[a] Nonsynonymous variant sites

[b] Synonymous variant sites

[c] Homoplasic variant sites (exclusively occurring in genes) relative to strains from the trachoma biovar.

[d] Number of insertion/deletion (indel) events among LGV strains.

[e] Due to alignments uncertainties concerning CT456/*tarp*, the effective number of homoplasic sites and indels cannot be established (they may range from to 5 to ~20).

We found that the whole genetic variability among LGV strains encompasses 1497 variant sites (1338 of them targeting genes and 159 occurring in IGRs) and about 100 indel events (equally distributed among coding regions and in IGRs) (Table 6.1). Although variant sites targeted more than half of the *C. trachomatis* genes, suggesting a random distribution of the mutations around the genome, we were able to detect several hypervariable genes among LGV strains (Figure 6.2). In fact, from the 448 *C. trachomatis* genes displaying variant sites among LGV strains (five out of the 453 genes only display indels), 44 genes (10%) and 14 genes (3%) carry at least five or 10 variant sites, respectively (Figure 6.2). Nevertheless, as homologous recombination is known to have occurred between strains from the LGV and trachoma biovars (37,175), we sought to stratify each variant site occurring in coding regions by distinguishing SNPs that become exclusively fixed in the LGV populations (non-homoplasic LGV-specific sites) from those that were likely imported from non-LGV strains through homologous recombination (homoplasic sites relative to trachoma biovar). Non-homoplasic LGV-specific sites likely result from specific selective pressures (such as those imposed in the complex bacteria-macrophages interactions or in the anorectal environment), whereas homoplasic sites are unlikely to play a role in the unique LGV phenotypes. Globally, of the 1338 sites occurring in coding regions that have a SNP among LGV strains, ~390 (~29 %) are homoplasic relative to strains from the trachoma biovar (Table 6.1 and Supplemental Table S6.2), which is in agreement with the homoplasy frequency observed (26%) within *C. trachomatis* species (37). However, it is worth noting that about 50% of these homoplasies occur in the typing gene CT681/*ompA*. Because of this, and due to the major impact of the *ompA* variability on the global trends, the results throughout the text are presented with and without *ompA* (Table 6.1). Moreover, six genes (CT050, CT051, CT652/*recD_2*, CT675/*karG*, CT679/*tsf*, and CT681/*ompA*; each one with more than ten homoplasic sites), corresponding to 1.3% of the genes with variability among LGV strains, account for 81% of homoplasies occurring in genes relative to the trachoma biovar (Figure 6.2 and Supplemental Table

S6.2). These genes could then be considered as the major targets of interbiovar recombination in *C. trachomatis*. When looking at the LGV-specific non-homoplasic sites (more than 900 sites), about 55% were nonsynonymous. These variant sites likely represent mutations that have become fixed as a result of the evolutionary adaptation of LGV strains to mononuclear phagocytes or the anorectal tract.

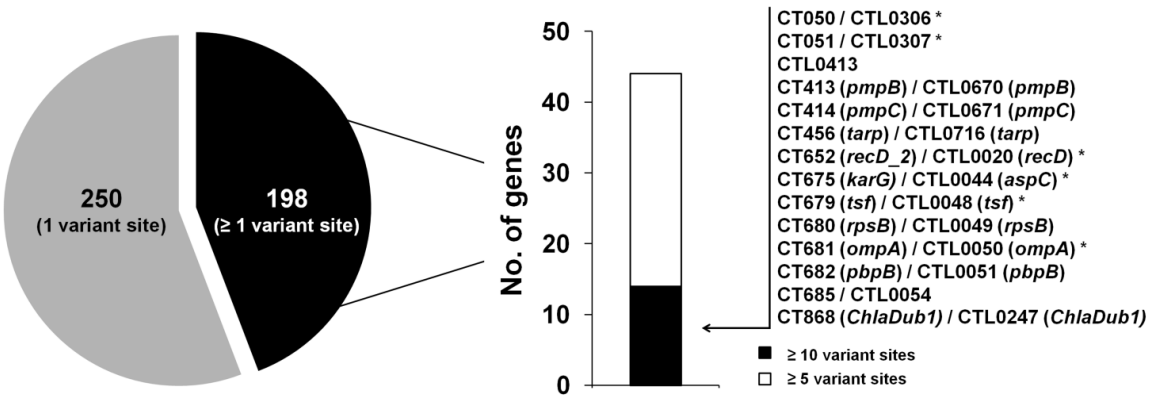


Figure 6.2. Distribution of the intra-LGV variant sites occurring in coding regions. The circular graph shows the number of *C. trachomatis* genes showing one (grey) or more than one (black) variant sites among LGV strains, whereas the bar graph highlights the number of genes displaying higher variability: at least five (white) or 10 (black) variant sites. The set of genes displaying hypervariability among LGV strains (at least 10 variant sites) are listed, where an asterisk (*) labels the ones with more than 10 homoplasic variant sites relative to strains from the trachoma biovar (major targets of interbiovar recombination). Due to alignments' uncertainties concerning CT456/*tarp*, the effective number of homoplasic sites cannot be determined, but its number expectedly ranges from to 5 to ~20. Thus, we used a conservative approach by not including CT456/*tarp* in the set of genes with more than 10 homoplasic sites.

6.4.2. Detection of genes more prone to accumulate mutations as a result of the LGV-specific infections

We analyzed the number of nonsynonymous and synonymous variant sites *per* gene in order to identify the genes that seem to have been favored to fix putative adaptive mutations. Figure 6.3 displays the total number of variant sites (and nonsynonymous variant sites) among LGV strains found for each *C. trachomatis* gene (from chromosome and plasmid). Globally, 714 nonsynonymous variant sites (607 if *ompA* is excluded) and 597 synonymous variant sites (486 w/o *ompA*), targeting 448 genes, were identified (Table 6.1 and Supplemental Table S6.2). The observation that about 45% of all variant sites in coding regions among LGV strains are synonymous along with the existence of 131 genes displaying a single synonymous variant site suggests the significant accumulation of neutral mutations by random genetic drift (248). In fact, according to the codon usage of the *C. trachomatis* (366), the chance of a random base substitution in a protein-coding region to be synonymous does not exceed 23% (data not shown), a value strikingly lower than that observed in this study. Also, as 109 genes were found to carry just one nonsynonymous variant site (Figure 6.3 and Supplemental Table

S6.2), it is rather plausible that several protein-altering mutations are not adaptive. In particular, this may reflect a scenario of Muller's ratchet (150-152,367) leading to the fixation of slightly deleterious mutations as a result of the genetic bottleneck derived from niche specialization and small population size of LGV strains. This kind of evolutionary signature was previously suggested to have also occurred upon both the *C. trachomatis* speciation (301) and the serovar radiation (173). Nevertheless, we were able to select 34 genes with a clear overrepresentation of nonsynonymous mutations among LGV strains, suggesting a role in pathoadaptation (Figure 6.3 and Supplementary Table S6.2). Remarkably, these genes (around 7.5% of all *C. trachomatis* genes with intra-LGV variability) account for 31.9% of all nonsynonymous variant sites detected among LGV strains and thus may be the most likely to be involved in distinct disease outcomes caused by LGV strains, ranging from a hemorrhagic proctitis to a suppurative lymphadenitis. In particular, 11 out of these genes (CT049, CT082, CT119/*incA*, CT223, CT233/*incC*, CT333/*uvrA*, CT442/*crpA*, CT456/*tarp*, CT711, CT868/*ChlaDub1* and CT875/*tepP*) displayed no silent mutations and at least 5 nonsynonymous variant sites, or a ratio of nonsynonymous/synonymous variant sites equal or above five (Figure 6.3 and Supplementary Table S6.2). Regarding the estimation of the *dN*/*dS* ratio, this evolutionary parameter was not very informative, as the majority of the genes exclusively showed nonsynonymous mutations hampering the application of the test. Considering that the vast majority of the 34 genes fall into specific categories with remarkable relevance in the *C. trachomatis* biology (167), the following sections will focus on the evolutionary trends of each gene/protein category regarding pathoadaptation of LGV strains. We will also highlight the genes carrying mutations that are specific of each LGV serovar, and emphasize the adaptive events that have been occurring throughout the clonal expansion of the epidemic L2b lineage.

6.4.3. Effectors and substrates of secretion systems

C. trachomatis is able to manipulate and subvert the host-cellular functions by translocating proteins (called "effectors") into the host cytosol. Among the multiple cellular processes known to be exploited by chlamydial effectors, which have been systematically revised during the last years (103,121,125-127,137), one can highlight apoptosis, cytokinesis, host cytoskeleton, nutrient transport, membrane trafficking pathways or immune responses. Hence, the regulation and subversion of these mechanisms are the hallmark behind the capacity of *C. trachomatis* to invade host cells and maintain its specialized intracellular replicative niche. With exception of the some prominent effectors (such as the proteases CPAF/CT858, CT441/Tsp, CT867/ChlaDub2 and CT868/ChlaDub1), the majority of the *C. trachomatis* effectors described so far were found to be translocated into the host cell cytosol by using a type III secretion (T3S) system (103,127,137,138). This system is also used by *C. trachomatis* to transport proteins to be localized into the inclusion membrane (Inc proteins, which given their singularity will be analyzed in a specific section) (129-132,324).

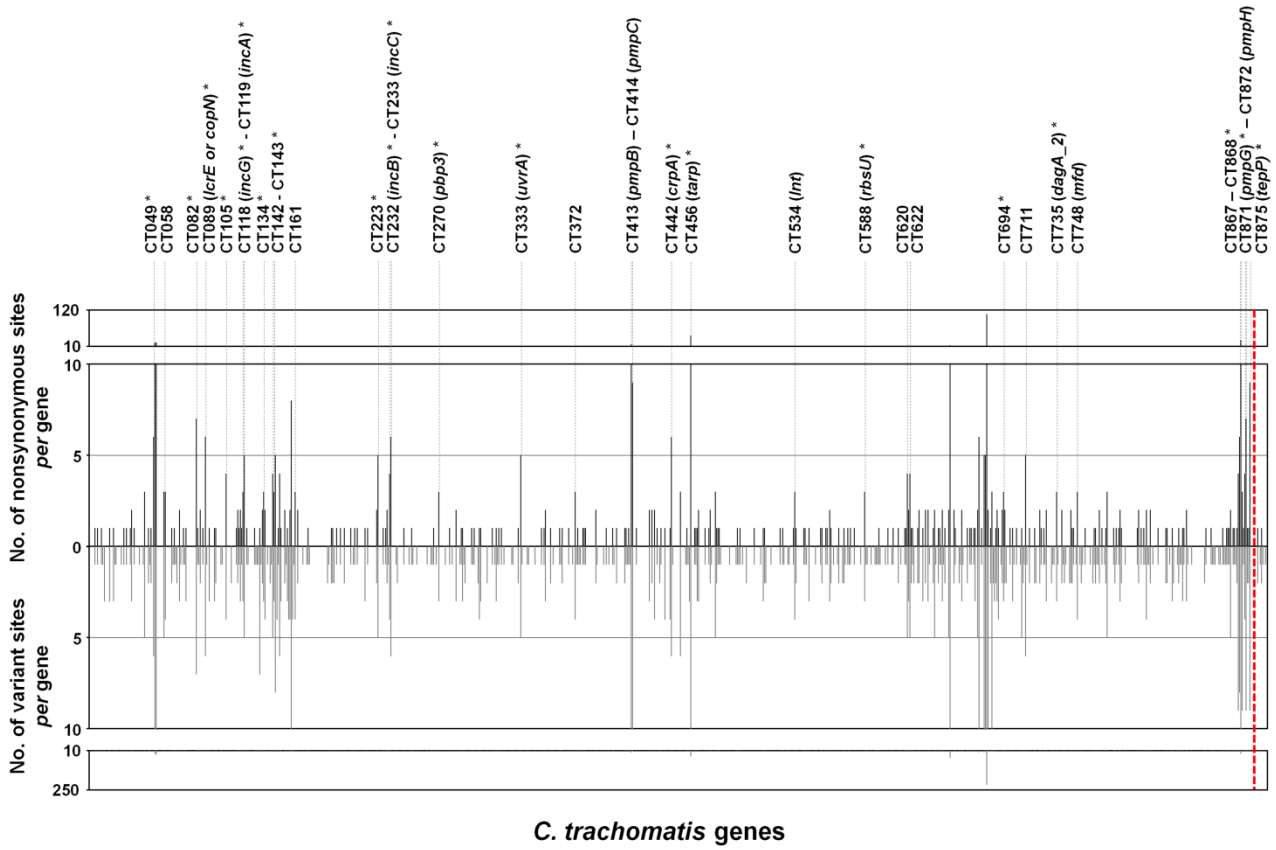


Figure 6.3. Hotspots of putatively adaptive intra-LGV mutations. The graph shows the total number of variant sites (lower bars) and the number of nonsynonymous variant sites (upper bars) among LGV strains detected for each *C. trachomatis* gene. The chromosomal genes are ordered according to the genome annotation of the D/UW-3/CX strain (GenBank accession number NC_000117). The plasmid genes (at the right side of the red dashed line) are ordered from the ORF1/*pgp7* to the ORF8/*pgp6*. The genes showing at least three nonsynonymous variant sites (and no synonymous variant sites) or a ratio of nonsynonymous/synonymous variant sites equal or above three are labeled above the graph. The asterisk (*) labels genes with *dN*/*dS* >1 and significant Z-test of positive selection or genes only with *dN* (for which the test cannot be applied). The CT089/*lcrE* (or *copN*) is the only gene from this set for which most of the nonsynonymous SNPs are homoplasic relative to non-LGV strains, making it less likely to be involved in variability of disease outcomes caused by distinct LGV strains.

Strikingly, we observed that 33 genes encoding effectors and/or T3S substrates (excluding Incs) (almost all of those described in the literature so far) (Supplemental Table S6.2) display variant sites among LGV strains, and the nonsynonymous variant sites are, in general, five-fold more frequent than the silent ones. In fact, these set of genes, accounting for just up to 5% of the *C. trachomatis* coding region, carry more than 20% of all nonsynonymous variant sites detected in this study (Fisher's exact test, $P < 10^{-7}$; odds ratio 4.88; 95% confidence interval, 4.05-5.87). Among the ones that mostly contribute to amino-acid alterations (Figure 6.3), we highlight the three early secreted T3S effectors,

CT456/Tarp, CT694 and CT875/TepP, pointed to play critical roles during the host-cell invasion and/or establishment of the nascent inclusions (103,121,126,368). Tarp (Translocated actin recruiting protein), which account for more than 40 amino acid alterations (only two silent changes were observed) among LGV strains, facilitates the invasion by promoting actin re-arrangements (140,141), whereas CT694 (possibly secreted at same time as Tarp) interacts with the cytoskeletal organizing protein AHNAK (266). In turn, the effector TepP (Translocated early phosphoprotein), which, similarly to Tarp, is tyrosine-phosphorylated upon bacterial entry, was recently found to act downstream of Tarp by inciting signaling cascades important for sustaining the *C. trachomatis* intracellular life-style (368). Other noteworthy examples are the putative effectors CT620, CT622 and CT711 (265,310) (Figure 6.3). CT620 and CT711 (like the genes CT619, CT621 and CT712, which also vary among LGV strains but in a less extent) encode proteins from a family of T3S substrates containing a domain of unknown function (DUF582) that are believed to target nuclear cell functions (310,317). The T3S effector and antigen CT622 (252,264,265) is highly variable among *C. trachomatis* strains (163,301), and is likely involved in pathogenic differences *in vivo*, as it belongs to a pool of genes implicated in virulence dissimilarities among two *C. trachomatis* ocular strains in nonhuman primates (164). Our analyses also showed that the genes CT867/*ChlaDub2* and CT868/*ChlaDub1*, which encode proteases possessing deubiquitinating (DUB) and deneddylating activities with a putative role in virulence (268), present a high number of intra-LGV variant sites. This is particularly relevant for the protease CT868/ChlaDub1, since it is targeted by more than 20 amino acid alterations and it plays a relevant role in subverting host cell functions by suppressing the NF-κB activation induced by several pro-inflammatory stimuli, which is an immune-evasion strategy applied by pathogenic organisms (269). Finally, we also observed that the recently identified T3S substrate CT105 (369) showed considerable variability among LGV strains. Nevertheless, we anticipate that its function is more likely associated with the phenotypic distinction between LGV and genital strains considering that: *i*) CT105 shows an overrepresentation of amino acid alterations between these two disease-groups (Chapter II) (301); *ii*) it is a pseudogene for ocular strains (301); *iii*) it shows higher expression for LGV strains (Chapter IV); and *iv*) there is LGV-specific 74-bp insertion within the putative promoter region of this gene that may justify the observed disease group-specific differential expression (Chapter IV). In conclusion, the activities involving subversion of the infected eukaryotic cells, in particular those mediated by the T3S system, are likely central also for the phenotypic variability displayed by LGV strains.

6.4.4. Inclusion membrane proteins (Incs)

The Inc proteins of *C. trachomatis* decorate the cytosolic face of the intracellular membrane-bound vacuole (inclusion) and have been found to subvert several host cell functions by establishing critical interactions with host proteins (255,293,294,307,308). In general, they are transported by a T3S mechanism and are inserted into the inclusion membrane likely due to the possession of an amino acid bilobed hydrophobic motif (129-132,324). We have previously shown that the majority of the known or putative Incs reveals a huge amount of amino acids changes that distinguish LGV strains from the remainder *C. trachomatis* strains (301,324,365), which suggests that this protein family plays

a relevant role in the unique tropism and invasiveness of LGV strains. In the present study, we observed that nonsynonymous variant sites targeting *inc* genes (about half of ~60 Incs described in the literature show intra-LGV variability) are four-fold more frequent than silent changes, which may underlie phenotypic dissimilarities within these strains. Putting together Incs and other T3S substrates/effectors (constituting ~10% of *C. trachomatis* coding region), they encompass 28.6% of all nonsynonymous variant sites (Fisher's exact test, $P < 10^{-7}$; odds ratio 3.68; 95% confidence interval, 3.12-4.35). Notably, nine in each ten variant amino acids in Incs are localized in regions of the proteins predicted to be on the cytoplasmic side of the inclusion membrane (Table 6.2 and Supplemental Table S6.2), supporting an adaptive role of these mutations as they alter the domains that may interact with host proteins. Importantly, this pool of *inc* genes: *i*) encode the most prominent Incs believed to play crucial roles in subverting the host cell machinery, namely CT115/IncD, CT118/IncG, CT119/IncA; CT222, CT223, CT229 and CT813 (255,260,294,308,370,371); *ii*) encode Incs implicated in the strong elicitation or modulation of host immune responses, such as CT147, CT232/IncB, CT233/IncC and CT442/CrpA (157,159,372,373); and *iii*) enroll the genes CT058, CT192 and CT214 that were found to display LGV-specific expression (324). Among Incs that most vary among LGV strains (Figure 6.3 and Table 6.2), we highlight CT119/IncA and CT223, which have SNARE-like (eukaryotic soluble N-ethylmaleimide-sensitive attachment protein receptors) motifs, and are involved in the subversion of the intracellular trafficking (255) by inducing the homotypic fusion of intracellular inclusions and protecting the inclusion from lysosomal fusion (for the former) (142-145) or by blocking host cell cytokinesis (for the latter) (260). In particular, the IncA has already been implicated in the disease severity, since nonfusogenic isolates were found to yield both less-frequent signs of infection and lower progeny (145,254,374). Another noteworthy example stands for the Inc CT442/CrpA (259), which is an antigen found to elicit specific CD8+ T cells responses (159). This may reflect a scenario of antigenic variation, although none of the intra-LGV variant sites fall into the described epitope ASFVNA(P)IYL ($CrpA_{63-71}$). Of note, these three Incs revealed 5-6 variant amino acids (and no silent mutations) among protein sequences from LGV strains (Table 6.2).

Table 6.2. Incs (or putative Incs) with variant amino acids among LGV strains.

Protein	length (aa) [a]	TM segments [a, b]	intra-LGV variant amino acids [c]
CT005 / CTL0260	363	[34-54] [60-81] [95-119] [124-146]	**I213T**
CT058 / CTL0314	367	[26-48] [53-76]	**P23S, A92S, P119S**
CT115 (IncD) / CTL0370 (IncD)	161	[51-75] [82-107]	**G158R**
CT116 (IncE) / CTL0371 (IncE)	132	[36-59] [64-87]	**K27E, E111G**
CT117 (IncF) / CTL0372 (IncF)	104	[38-62] [70-91]	V87I
CT118 (IncG) / CTL0373 (IncG)	167	[33-57] [63-88]	**A32T, E122K, F141L**
CT119 (IncA) / CTL0374 (IncA) [d]	273	[35-59] [64-84]	Q64H, **G180S, R200C, Q206R, S267L**
CT134 / CTL0389	137	[80-98] [105-122]	**A20V, A42V,** F98S
CT135 / CTL0390	360	[210-236] [242-268]	**M156I, I342M**
CT147 / CTL0402	1449	[79-99] [105-124] & [849-870] [876-896]	**F9S, K388Q, K1061E, V1422L(M)**
CT192 / CTL0444	232	[32-55] [61-82]	**Q109L**

CT195 / CTL0447	363	[53-74] [81-104] & [173-194] [196-215] & [238-252] [254-268]	**P48L**
CT214 / CTL0466	547	[36-58] [64-89]	**R385S**
CT222 / CTL0475	128	[39-63] [69-92]	**R36C, A102S**
CT223 / CTL0476	268	[38-61] [67-91]	**F32S, A103D, C133Y, E144D, C219S**
CT227 / CTL0479	133	[36-64] [66-88]	**S2A**
CT229 / CTL0481	215	[42-65] [71-90]	**R101C**
CT232 (IncB) / CTL0484 (IncB)	115	[31-61] [67-90]	**S15R,** I34V, I77V**, A97T**
CT233 (IncC) / CTL0485 (IncC)	178	[100-127] [140-164]	**M5I, S6P, D7G, I8V, K11R, I14V**
CT249 / CTL0500A	116	[51-72] [78-97]	Y51C
CT288 / CTL0540	564	[36-58] [65-88] & [242-263] [269-291]	**A109T**
CT442 (CrpA) / CTL0701 (CrpA)	150	[40-62] [68-88]	**A24V, K29Q,** I48M, **I104V, D133G, V137A(E)**
CT618 / CTL0882	266	[213-236] [242-262]	**K78Q**
CT813 / CTL0184	264	[41-61] [68-94]	**T189A**

[a] Based on the protein sequence annotation of L2/434/Bu (GenBank accession number NC_010287).

[b] Positions of transmembrane (TM) domains in the corresponding Inc proteins, obtained from literature (129), except for CT223, for which we previously found only the two identical TM domains (324). The positions from the D/UW-3/CX strain [used by Dehoux et al (129)] were converted to the corresponding positions in the L2/434 protein annotations. Dehoux and colleagues (129) also described the TM domais of the putative Inc CT873, whose putative homolog protein (based on the genome location) in the L2/434/Bu strain (CTL0252) display one amino acid variant among LGV strains. However, possibly as a result of mis-annotation, the annotated proteins are radically different or truncated in one or both strains (46), which hampers the transposition of the TM location.

[c] Intra-LGV variant residues in regions predicted to be on the cytoplasmic side of the inclusion membrane are in bold (we considered that the loop region within two TM segments faces the lumen of the inclusion). The changes are relative to the amino acid in the same position in L2/434/Bu proteins.

[d] Taking into account the putative misleading annotation of the CT119/IncA protein in the L2/434 annotation, where it is annotated with 246 aa [i.e., 27 aa shorter than expected at N-terminal (143)], we opted to present the protein length (273 aa) as annotated for the D/UW-3/CX strain (NC_000117).

6.4.5. Polymorphic membrane proteins (Pmps)

All nine *pmp* genes accumulated mutations among LGV strains, although we have observed distinct dynamics of mutational fixation among them. In fact, while CT413/*pmpB*, CT414/*pmpC,* CT871/*pmpG* and CT872/*pmpH* over-accumulated nonsynonymous substitutions (about four-fold more frequent than silent mutations) that result in protein variation among LGV strains (Figure 6.3 and Supplementary Table S6.2), the majority of polymorphisms within the remaining five *pmp* genes cause synonymous changes, indicating some degree of purifying selection for the encoded Pmp to preserve its structure and function. These autotransporter proteins are relevant *C. trachomatis* antigens (110,165,250,252,277,278,375) and adhesins (109-112), a function known to display tissue specificity (109). Interestingly, a recent study (376) showed some evidence that the mutational pattern of most Pmps likely underlie the efficiency of attachment to host cells in absence of centrifugation in culture, which is a historical differentiating phenotype between LGV strains (high attachment efficiency) and non-LGV strains (low attachment efficiency) (377). Considering the existence of specific molecular

signatures for Pmps, we sought to perform a detailed analysis of the location of all variant amino acids. For CT414/PmpC, we observed that one variant amino acid affects one GGA(I,L,V) motif (positions 541-544), although the motif is not lost as L1/115 and L1/1322/p2 have a GGAV motif whereas the remaining strains display GGAI. Also, we have previously reported (169) that a different clone of the L1/440 strain possessed an extra GGAI motif when comparing with the remainder LGV strains. This tetrapeptide motif, as well as the FxxN motif, is believed to be essential for the adhesion function of *Chlamydia* Pmps (378), and their number and distribution along the proteins is known to vary among Pmps and also among same-species strains for the same Pmp (109,111,112). Additionally, we found one amino acid difference immediately downstream of one FxxN motif in CT862/PmpH and another adjacent to one GGAI motif in CT874/PmpI, which may somehow promote some structural or functional constraints regarding the adhesion activity. Finally, five *pmp* genes (CT413/*pmpB*, CT414/*pmpC*, CT869/*pmpE*, CT870/*pmpF* and CT872/*pmpH*) display mutations specific of strains from a particular LGV serovar (see section “Mutations targeting specific LGV serovars” for details). Overall, the recognized importance of the Pmp family in the *C. trachomatis* biology and diversity (109,167) along with the sequence variation herein highlighted might help us infer that Pmp-mediated specific adhesion properties or differential immune evasion strategies may also account for phenotypic dissimilarities among LGV strains.

6.4.6. Plasmid and plasmid-related chromosomal genes

Efforts to dissect the molecular features behind the *C. trachomatis* phenotypes have been historically focused on the analysis of the chromosomal variability among strains [reviewed by Nunes et al. (167)]. However, increasing data supports that the naturally occurring ~7,5kb *C. trachomatis* plasmid plays an important role in the pathogenesis of infection and disease (203,207,210,337,379-383), and thus, it may also be a key determinant of the dissimilar tissue tropism, virulence and ecological success among strains. Indeed, although the *C. trachomatis* plasmid shows low polymorphism among strains, its phylogenic profile still reveals a strict co-evolution with the chromosome by displaying a parallel pattern of segregation of strains regarding their disease outcomes (37,166), and its transcriptional dynamics somehow correlate with the tropism of the strains (355). Moreover, the exchange of plasmids between *C. trachomatis* strains is thought to be rare (37), and recent plasmid-mediated transformation assays suggest that the plasmid ability to replicate may be biovar-specific, or even strain-specific (384,385), corroborating that its biological role may underlie strain-specific phenotypes. Finally, as the plasmid is a primary transcriptional regulator (207,379,383) of multiple chromosomal virulence-associated genes [a function recently attributed to the plasmid protein ORF6/Pgp4 (207)], the molecular mechanisms behind the putative plasmid-mediated phenotypic dissimilarities among strains may be highly diverse, and may rely on the genetic variability displayed by its chromosomal gene targets. On behalf of this, we also looked at the genetic variability among LGV strains targeting both the plasmid (also reported by other authors) and the plasmid-related chromosomal genes (this study). Only seven variant sites were observed in plasmid genes (only two leading to an amino acid replacement), and three indel events, two of them occurring in the

gene ORF1/*pgp7* (Supplementary Table S6.2). One of these events is a single base deletion that prematurely leads to the loss of the 45 amino acids from the C-terminus of the predicted protein in the L2/434 strain (386), whereas the other is an 85-bp frameshift deletion that likely inactivates the gene in the serovar L3 strain (L3/404) (37). Curiously, the gene ORF1/*pgp7* was previously shown to be both non-essential for the *C. trachomatis* growth *in vitro* (207,380,382) and to be also interrupted in the plasmid of other naturally occurring *Chlamydiae* (387), which corroborates its species- or strain-specific dispensability, launching the hypothesis that this protein may play a role on some very specific phenotypic specificities including those displayed by different LGV strains (e.g., all L2b strains display a non-truncated protein). The third indel (27 bp deletion) occurs in the IGR upstream of the ORF1/*pgp7* solely for the strain L2b/UCH-2. Nevertheless, this event should be eyed with cautious as it involves a 4 x 22 bp repeated region (in the origin of replication) that may yield assembly artifacts. Of note, the transcriptional regulator Pgp4 (207), which displays a LGV-specific amino acid at position 5, is 100% conserved among LGV strains. Moreover, whereas sRNA-2, the highly expressed small anti-sense RNA (anti-sense to ORF2/*pgp8*) (355), shows no variability, the sRNA-7 (anti-sense to ORF7/*pgp5*) carries one mutation in the strain L1/1322/p2.

Regarding the 32 virulence-associated chromosomal genes regulated by Pgp4 (207), we found that 20 are polymorphic among LGV isolates (Supplementary Table S6.2), and from the 124 intra-LGV variant sites found in this pool of genes, two thirds (82/124) altered the corresponding protein. Among the most variable genes, we highlight two sets of contiguous genes: the paralogously related CT049, CT050 and CT051, and the putative T3S substrates-encoding CT142, CT143 and CT144 (369). The first set encodes three Pmp-like proteins and putative antigens (264) that are among the most polymorphic proteins in *C. trachomatis* (163), and for which the high variability have been suggested to underlie mechanisms of immune evasion (257). However, despite CT049 and CT050 have been already found to localize to the inclusion lumen and to be associated with the inclusion membrane (257), their relevance for chlamydial pathogenesis or phenotypic variability is still to be clarified. CT050 and CT051 carry several homoplasies relative to trachoma biovar strains (Figure 6.2), but their high variability among LGV strains still account for a high number of non-homoplasic intra-LGV variant sites. Thus, they could serve as potential targets of positive selection driving pathogenic diversity among LGV strains, as previously suggested (301). Regarding the CT142-CT144 genes [which together with CT049 and CT798/*glgA* are the most significantly transcriptionally regulated genes by the virulence plasmid (207)], they encode T3S substrates that possibly act coordinately in late or early stages of the *C. trachomatis* life cycle (369). CT142 and CT143 were already identified as effectors (369), whereas CT144 have been found to be an antigen capable of eliciting protection against vaginal challenge with *C. trachomatis* in mice (158), and to carry polymorphisms associated with rectal tropism in serovar G isolates (172). Considering the latter observation, we investigated whether the rectal L2b strains shared any exclusive amino acid with those serovar G rectal strains in CT144, but no correlation could be established. Finally, we also identified the variant sites differentiating the plasmid-bearing strain (L2/434/Bu) from the same-serovar naturally occurring plasmidless strain (L2/25667R), and found some discrepancies regarding the list of variant sites previous found in a L2-*versus*-L2R microarray-based whole-genome comparison (379).

We have now identified one indel event [a 3bp deletion in the gene CT534/CTL0796 (*Int*) in the plasmidless strain] and 8 SNPs (Supplementary Table S6.2). Six SNPs are nonsynonymous, and three of them target the gene CT082, which encodes a T3S substrate (312,313) that likely remains associated with the chlamydial cell surface at late stages of the life cycle (313). Considering that this gene is one of the very few whose transcription is repressed by the plasmid (207) and that plasmidless viable strains are rarely isolated (208-210), one could wonder about the essentiality of CT082 in *C. trachomatis* biology. On the other hand, as we have already detected substantial transcript' levels for CT082 in several LGV plasmid-bearing strains (data not published), and taking into account that one altered amino acid is shared by one proctitis-causing L2b strain (L2b/Ams3), and that another occurs in a protein position where four different amino acids are observed among LGV strains (highly unlikely to occur by change in a protein with 560 amino acids), this speculation is questionable. Additionally, one of the others L2-versus-L2R nonsynonymous mutations occur in the *inc* CT058, which we previously shown to have LGV-specific gene expression and to be a pseudogene for most ocular strains (324).

6.4.7. Mutations targeting specific LGV serovars

Although the LGV strains are able to cause a wide range of clinical manifestations, nearly all cases of the anorectal syndrome with acute proctitis (current epidemic atypical LGV presentation) are caused by L2b strains (46,362,388-390), whereas the remainder LGV serovars are essentially associated with the inguinal syndrome with painful inguinal lymphadenopathy (typical bubonic LGV presentation) (37,46,358,391). Intriguingly, previous comparative genomic analyses involving both L2 and L2b strains showed no additional coding capacities that may justify the differences in clinical manifestations (46) and that the epidemic outbreak of LGV is certainly a result of clonal expansion of a serovar L2b strain (37). As the genome-based species phylogeny showed that the main genetic variability within the *C. trachomatis* species is defined by the different serovars (37,365), one could expect that some serovar-specific mutational signatures may underlie the distinct phenotypes described above.

Hence, we searched for the variant sites/indels among LGV strains that remained serovar-specific (even after the action of recombination). We detected 33 serovar-specific indels (Table 6.1 and Supplementary Table S6.2), where 20 occur in IGRs and involve just 1 to 3 bp. From these, we highlight the indel of 1-3 bp in the IGR upstream from CT042/*glgX*, which perfectly discriminates each of the four LGV serovars according to the number of G nucleotides (six to nine) in a homopolymeric tract. Although we cannot infer about the impact of such variability, homopolymeric tracts are known to have a major role in phase variation of bacteria, which is a mechanism that creates phenotypic diversity, such as the differential affinity of pathogens for different host anatomical niches or expression of virulence phenotypes (392-395). Interestingly, a poly(G) tract has been described to mediate phase variation in *pmp10* of *C. pneumoniae* (396,397). If that would be the case for the detected indel, it could be involved in the expression regulation of the gene CT042/*glgX*, which encodes a glycogen-degrading enzyme (glycogen hydrolase) involved in the glycogen metabolism. It

has been suggested that glycogen may be associated with niche adaptation, since it may be advantageous for chlamydial survival in mucosal tissues by acting as an essential carbon source within nutritionally limiting inflammatory environments (207). Whether this confers any advantage for colonizing the rectal mucosa by L2b strains cannot be determined at this time.

We also identified 557 serovar-specific variant sites, where 2.5 %, 23.9 %, 56.0 % and 17.6 % of them are serovar L1-, L2-, L3- and L2b-specific, respectively (Table 6.1, Figure 6.4 and Supplementary Table S6.2). An important bias associated with this analysis of serovar-specificity relies on the availability of a low number of strains representing some of the LGV serovars. Furthermore, whereas the use of the branch-site test of positive selection (241,242) restricted to the tree of LGV strains would be an useful approach for associating specific mutations with phenotypes (301), its application at serovar level would mostly be a confounding factor rather than informative as some LGV serovars are dispersed in several tree branches and share the same phenotype (the typical bubonic LGV). The scarce number of serovar L1-specific variant sites can be essentially explained by the fact that two South Africa strains (L1/115 and L1/224) have been shown to display a high number of homoplasies and a rather unusual L1/L2 recombinant *ompA* genotype (37). Considering this pattern for L1 strains and the fact that only two L2 and one L3 strains were fully-sequenced so far, we opted not to elaborate assumptions regarding the variant sites (and their gene targets) specific of these serovars because they may be strain- (not serovar-) specific. Therefore, we will thus focus on the L2b strains. Moreover, when looking at the impact of the serovar-specific mutations on the protein sequence, only the L2b-specific mutations are overrepresented by amino acid changes (Figure 6.4). The L2b-specific variant sites are dispersed throughout the chromosome (similarly to the other LGV serovars), but only 10 loci display more than one variant site. These set of loci includes genes encoding proteins whose functions may hypothetically contribute to the atypical clinical manifestations displayed by the epidemic L2b strains, such as CT118/IncG, CT333/UvrA, CT694, CT748/Mfd and CT872/PmpH (each one carrying no silent mutations and two or more nonsynonymous L2b-specific mutations, where CT872/*pmpH* carries a unique 9bp L2b-specific insertion). Two of these proteins (IncG and CT694) are known to directly interact with the host-cell. The IncG protein colocalize with kinases of the Src family (258) and interacts with the host protein 14-3-3β preventing initiation of apoptosis by the host cell (146,308,371), whereas the immunodominant antigen (157) and T3S effector CT694 has been implicated in the remodeling of actin filaments during invasion or early stages of *C. trachomatis* infection through the interaction with the human AHNAK (266,267). The genes CT333/*uvrA* and CT748/*mfd* are functional partners as both encode proteins involved in the nucleotide excision repair. It is known that bacterial pathogens are constantly challenged by DNA-damaging agents (namely reactive oxygen and nitrogen species) produced by host phagocytes, and that, for instance, the *uvrA* gene is up-regulated in *Mycobacterium tuberculosis* grown in human macrophages (398). Considering that both genes show variability between all LGV serovars (Figure 6.3), and that more than half of the nonsynonymous variant sites are L2b-specific, we wonder if this specificity does not somehow impact their ability to infect macrophages and proliferate to lymph nodes, in contrast with the strains from the remainder LGV serovars. CT872/PmpH, as described in a previous section, is a putative adhesin (112) and/or antigen (250,252). Curiously, a recent study (399)

took advantage of the CT872/*pmpH* genetic variability among LGV strains (namely, of the L2b-serovar-specific 9-bp insertion) to construct a real-time PCR primer/probe detection set reliable for identification of the L2b variant. In a similar fashion, all loci here detected to carry serovar-specific variant sites/indels are potential targets for the development of new molecular diagnostic tests aiming to differentiate LGV strains, in particular, the epidemic L2b strains. Noteworthy, we verified that the presence of shared polymorphisms between the recombinant L1 strains and L2b strains clearly masks numerous L2b-specific mutations. In fact, by removing these "outliers" L1 strains, we would obtain at least 50 additional L2b-specific variant sites (data not shown).

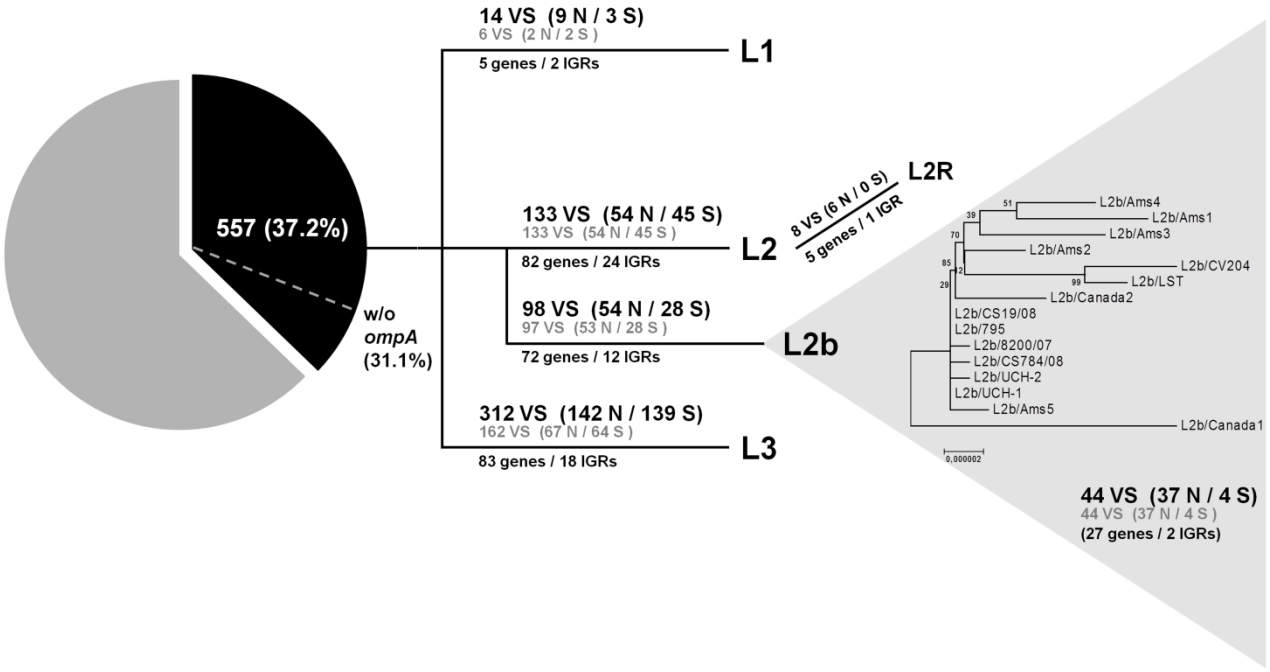


Figure 6.4. Mutations targeting specific LGV serovars and underlying intra-L2b variability. The circular graph (in black) shows the proportion of serovar-specific variant sites (i.e., variant sites distinguishing a singular serovar from the remainder LGV serovars). The dashed line reflects that proportion when the typing gene CT681/*ompA* is excluded from the analysis. The total number of variant sites (VS) as well as the ratio of nonsynonymous (N) and synonymous sites (S) is depicted in each branch of the virtual tree. Results are shown with (black) and without *ompA* (grey). The number of genes and IGRs harboring the variant sites is also presented below each branch. It is also displayed the number of sites differentiating the plasmid-bearing strain L2/434/Bu (L2) from the same-serovar plasmidless strain L2/25667R (L2R). The grey triangle highlights the genetic variability among the L2b isolates as a result of the fixation of mutations throughout the clonal expansion of the epidemic L2b lineage. This phylogenetic reconstruction involves the complete genome sequences of 15 L2b strains isolated in the last decade in North America and Europe, and was performed by using the neighbor-joining method with bootstrapping (1000 replicates) based on distance estimates using a Kimura two-parameter model for substitution events.

6.4.8. Mutational events that occurred throughout the clonal expansion of the epidemic L2b lineage

It is now well-established that the L2b variant has been circulating in the human population for a long-time, and that the proliferation of the L2b infections causing mostly acute proctitis (anorectal syndrome) among MSM in North America and Western Europe is a result of the expansion and transmission of that "old" variant clone (37,46,66,388,400). This proliferation has been suggested to be a function of the increase of both the network of sexual contacts among MSM and molecular screening and typing of associated infections (5,67,388,400,401). Nevertheless, despite the high degree of clonality of L2b strains (37), there are some reports of L2b infection causing either unusual disease outcomes (such as arthritis, pharyngeal syndrome or persistent genital ulcer) (66,67,69,77,402) or proctitis accompanied by inguinal lymphadenopathy (66,68,71,403-405) suggesting some degree of pathoadaptation and phenotypic variability even among L2b strains. In this regard, we also aimed to investigate which mutations were fixed and which genes were targeted by selection throughout the clonal expansion of the L2b lineage, as it constitutes a rare opportunity to track the adaptive dynamics underlying the microbe-host "arms race" that takes place essentially during infection of the rectal mucosa. This analysis involved 15 L2b strains isolated in the last decade in North America and Europe (Supplementary Table S6.1), and includes two additional newly fully-sequenced L2b strains from Portugal.

Globally, we detected 44 variant sites and 7 indels among L2b strains (Table 6.1, Figure 6.4 and Supplementary Table S6.2). Remarkably, 90.2% (37/41) of the variant sites occurring in coding regions involve amino acid replacements, which clearly points to a mutational scenario indicative of adaptation. The set of genes targeted by nonsynonymous mutations includes genes encoding: *i*) antigens (e.g., CT316/RplL and CT743/HctA) (252,264); *ii*) proteins likely implicated in virulence and/or subversion of host-cell functions (CT082, CT135, CT172/TrpA, CT203, CT301/PknD, and CT456/Tarp) (51,140,156,180,202,207,406); *iii*) proteins involved in the control of transcription (CT588/RbsU and CT636/GreA) (115); and *iv*) proteins with relevant secretory/translocation/transport functions (CT448/SecF, CT510/SecY and CT580) (407,408). Nevertheless, it is worth noting that SNPs occurring in most of these genes affect few strains. This is the case of CT743/*hctA*, encoding a histone like protein involved in the conversion of *C. trachomatis* reticulate bodies to elementary bodies late in the life-cycle (117), which presents two non-silent mutations for the strain L2b/Canada1. Also, CT510/SecY, which mediates translocation of the virulence chlamydial protease/proteasome-like activity factor (CPAF) (408), harbors a non-silent mutation for a single isolate from Amsterdam (L2b/Ams1). Of note, a homologue of the CT510/SecY in the sec-dependent pathway, the protein CT448/SecF, also carries an amino acid alteration for another isolate from Amsterdam (L2b/Ams5).

CT456/*tarp* is unequivocally the one with higher variability among L2b strain showing nine non-silent variant sites (and no silent mutations) and two indels. We detected an in-frame 147-bp deletion solely for the strain L2b/Ams3 (isolated in 2004 in Amsterdam, Netherlands). Interestingly, BLAST analysis against all *tarp* sequences available in GenBank revealed that this deletion is shared by two rectal isolates from North America: one serovar L2 strain (L2/98, from Ottawa, Canada) and

one serovar L2b strain (L2b/102, from San Francisco, USA) (287). Although no information is available regarding the year of isolation and clinical or sequence data from these isolates, this scenario strongly suggests transatlantic transmission of this particular L2b variant, and may corroborate the assumption that the L2b variants causing the current "outbreak" among MSM in Europe were imported from the MSM population in the USA in the 1980s by the end of the last century (66,388,390,400). It is also noteworthy that the 147-bp deletion affects the number of tyrosine-rich repeat regions of Tarp N-terminal, where these strains have five regions whereas the remaining LGV strains with rectal tropism display six (287). Considering that: *i*) the non-LGV *C. trachomatis* strains with tropism to epithelial cells of the conjunctiva or genital mucosa (even the ones isolated from the rectum) have an even lower number of tyrosine-rich repeats in Tarp (one to four) (287); *ii*) L1 and L3 strains (associated with the typical bubonic LGV) possess the highest number (nine) (287); *iii*) it is believed that enhanced tyrosine phosphorylation of Tarp N-terminal is likely correlated with the strength of interactions established with proteins from immune signaling pathways (276); and *iv*) it is hypothesized that the ability of the LGV strains to cross the mucosa epithelium and to infect mononuclear phagocytes may rely on the increased tyrosine phosphorylation upon host cell entry (167,276); it thus seems that although the requirements for Tarp tyrosine phosphorylation for the colonization of the rectal mucosa are low, the intermediate number of tyrosine-rich repeat regions evidenced by L2b strains likely provide them with capability for dual tropism (some L2b strains were already associated with bubonic disease) (66,68,71). Nevertheless, it could be interesting to perform a Tarp-based typing large-scale epidemiological survey on rectal strains to evaluate the frequency of those repeats in L2b isolates.

Considering what we observed for the above described deletion event, we further looked at other mutations potentially showing geographical linkage. The more relevant finding concerns the identification of two L2b strains (L2b/CV204 and L2b/LST) isolated in France segregated in the same clade (Figure 6.4) as they are the only L2b strains with five specific SNPs and one 12-bp insertion in CT456/*tarp* (leading to an additional motif NIYE in the tyrosine-rich repeat region of the encoding protein) and two SNPs in the genes CT195 and CT588/*rbsU.* Furthermore, their genomes differ solely by 7 SNPs, which clearly suggests that although these strains were isolated in different years (2006 and 2008) and cities (Paris and Tourcoing), they are likely related.

Although all L2b isolates here analyzed had identical *ompA* sequences, it is noteworthy that this typing gene, which encodes the major outer membrane protein (MOMP), has also been detected to vary among L2b isolates, with most mutations affecting the surface-exposed variable domains of the protein, which are expected to be under the selection of the host-immune system (72,388,409). Nevertheless, this scenario was found to occur worldwide across multiple *C. trachomatis* serovars(63), so no speculation can be made about the specific role of MOMP in the pathoadaptation of L2b strains. Finally, a previous study using multilocus sequence typing over several L2b strains collected worldwide revealed that all analyzed strains isolated since the 2003 epidemics contain one of the two alleles of the gene CT046/*hctB* identified among the strains that were isolated in the 80s in San Francisco, USA (388). Considering that all strains analyzed in our study (including the two recently sequenced strains from Portugal) have that same CT046/*hctB* allele, this corroborates the assumption of the existence of a single source of origin for the current epidemic LGV worldwide.

6.4.9. Variability among LGV strains involving non-coding regions

The IGRs of bacterial genomes encompass promoter regions, Shine-Dalgarno ribosome binding sequences or small noncoding RNAs, which make these regions targets of selection (410-413). Thus, we have also inspected these regions to search for genetic variability among LGV strains, as mutations in IGRs may not be innocuous but compromise the gene regulation. In fact, several reports in bacteria, including *C. trachomatis*, showed that few SNPs or indels in non-coding regions may be enough to impact transcriptional or translation stages, probably because the specificity of both the transcription factors and the RNA polymerase can be nucleotide-specific, and regulatory small noncoding RNAs can discriminate mRNA regions that differ by a single nucleotide (303,410,414-416).

Eight IGRs were found to display three or more variant sites distinguishing LGV strains (Supplementary Table S6.2). In particular, three of them present five or more variant sites: the IGR CT169/*trpR*-CT170/*trpB*, the IGR CT231-CT232/*incB*, and the IGR CT680/*rpsB*-CT681/*ompA*. Regarding the latter, several mutations detected are homoplasic relative to non-LGV strains and they fall into a region known to be a hotspot of homologous recombination in *C. trachomatis* genomes (37,170,365). Therefore, the impact of such variant sites lacking LGV specificity on phenotypic dissimilarities is implausible. The IGR CT169/*trpR*-CT170/*trpB* belongs to the *trpRBA* operon (417), which includes the tryptophan operon repressor gene (*trpR*) and the genes encoding the tryptophan synthase (*trpA* and *trpB*). The full activity of *trpRBA* is directly correlated to the ability of the *C. trachomatis* to respond and resist to the protective effect of the host interferon-gamma (IFN-γ) (which limits the tryptophan availability), and is mandatory to any *C. trachomatis* strain to infect the genitalia (51). Moreover, this mechanism may contribute for the unique tropism and inflammatory outcomes of *C. trachomatis* LGV strains, since they were found to be more resistant to the inhibitory effect of IFN-γ than the non-LGV genitalia-infecting strains (51). In this perspective, as a 10bp-insertion (solely for L3/404) and four out of the six variant sites fall into the *trpBA* operator sequence (202,418), we wonder if this genetic variability contributes for phenotypic dissimilarities among LGV strains. For the IGR CT231-CT232/*incB*, all variant sites are L2-specific and fall into a region located 33 bp upstream from the translational origin of the gene encoding the IncB, a gene that in turn also presents four L2-specific nonsynonymous mutations. Furthermore, the gene CT232/*incB* likely forms an operon with the genes CT233/*incC* and CT234 (102,115,261,324), so that the mutations in the IGR CT231-CT232/*incB* may also impact the gene expression of the latter. It is noteworthy that the upstream IGR and two genes CT232/*incB* and CT233/*incC* account for 15 variant sites among LGV strains. Also, CT233/*incC* displays an overrepresentation of nonsynonymous mutations among LGV strains in the 5'-end, as previously noticed (301), where the secretion signal necessary for its type III secretion is encoded (130,132,324). Considering these findings and the involvement of IncB and IncC in the modulation of the host immune responses, likely contributing to the immunopathology associated with chlamydial infections (372,373), it is tempting to speculate that the polymorphism within this region may underlie distinct disease outcomes caused by LGV strains. Finally, there is another IGR that is worth mentioning, the IGR CT683-CT684, as it shows four L2b-specific variant sites, a pattern

followed by the downstream gene CT684 (encodes an ABC transporter associated protein) that display two L2b-specific variant sites (one silent and one non-silent SNP). These are two examples where the selection seems to have been similarly targeting the whole transcriptional unit, but many more may exist. In this regard, we believe that it could be interesting to perform a whole-genome survey extended to multiple *C. trachomatis* strains displaying differential tropism (ocular tract, genitalia and mononuclear phagocytes). In fact, the existence of *C. trachomatis* operons where both the coding and the regulatory regions display a mutational pattern that is specific of a single phenotype (resulting from a directional selective pressure) may constitute interesting targets for functional studies.

6.5. Concluding remarks

While early diagnosis is crucial to avoid clinical complications and to decrease further transmission of LGV strains, it seems clear that additional and systematic strategies of genomic characterization of the infecting strains may considerably enhance our knowledge regarding the pathobiology of LGV strains. Therefore, the deep analysis described here of the genetic variability among LGV isolates has allowed us to determine an overrepresentation of amino acid changes targeting pathogen-host interacting proteins and other virulence determinants, namely effectors and/or T3S substrates, Incs and plasmid-regulated virulence factors. We also found SNPs and indels differentiating strains causing specific LGV disease outcomes, such as the typical bubonic LGV and the anorectal syndrome. In this regard, it is tempting to speculate that L2b-specific mutations are advantageous as revealed by the higher clinical prevalence of L2b strains in the worldwide LGV epidemiological picture (identified in recent bubonic LGV cases and in nearly all cases of LGV-proctitis). In addition, our discovery of dozens of serovar-specific variant sites may be the basis for the future development of discriminatory diagnostic/typing methodologies. By evaluating the evolutionary pathway beyond the L2b clonal radiation, we were also able to track the adaptive dynamics underlying the microbe-host “arms race” that takes place during infection of the rectal mucosa. Finally, the recent advance of transformation and gene knockout experiments in *C. trachomatis* (206,213,216,217) may constitute the basis for functional genomics studies aiming to dissect the impact of the identified polymorphisms on the phenotypic dissimilarities (such as virulence or tropism) and pathobiology of LGV strains. Ultimately, it may hopefully open new opportunities for developing prophylactic or therapeutic strategies against *C. trachomatis* LGV infections.

Acknowledgements

We thank Luis Vieira and Daniel A. Sampaio from the Sequencing core facility at the Portuguese National Institute of Health for their excellent technical assistance. VB was recipient of a Ph.D. fellowship (SFRH/BD/68527/2010) from the Fundação para a Ciência e a Tecnologia.

Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics

Manuscript with minor changes published in

2013, Infection, Genetics and Evolution, 17:23-32.
Borges V., Ferreira R., Nunes A., Sousa-Uva M., Abreu M., Borrego M.J., Gomes J.P.
Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics.
(http://www.ncbi.nlm.nih.gov/pubmed/23542454)

These results were presented as a poster in the Sixth Biennial Meeting of the *Chlamydia* Basic Research Society, San Antonio (Texas, USA), March 19-22, 2015.

Personal contribution

VB contributed to the design of the study, performed a significant amount of the experimental work, interpreted data and wrote the manuscript.

7. Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics

7.1. Abstract

It is assumed that bacterial strains maintained in the laboratory for long time shape their genome in a different fashion from the nature-circulating strains. Here, we analyzed the impact of long-term *in vitro* propagation on the genome of the obligate intracellular pathogen *Chlamydia trachomatis*. We fully-sequenced the genome of a historical prototype strain (L2/434/Bu) and a clinical isolate (E/CS88), before and after one-year of serial *in vitro* passaging (up to 3500 bacterial generations). We observed a slow adaptation of *C. trachomatis* to the *in vitro* environment, which was essentially governed by four mutations for L2/434/Bu and solely one mutation for E/CS88, corresponding to estimated mutation rates from 3.84×10^{-10} to 1.10×10^{-9} mutations *per* base pair *per* generation. In a speculative basis, the mutations likely conferred selective advantage as: *i*) mathematical modeling showed that selective advantage is mandatory for frequency increase of a mutated clone; *ii*) transversions and non-synonymous mutations were overrepresented; *iii*) two non-synonymous mutations affected the genes CTL0084 and CTL0610, encoding a putative transferase and a transcriptional regulatory protein, respectively, which are families known to be highly prone to undergone laboratory-derived advantageous mutations in other bacteria; and *iv*) the mutation for E/CS88 is located likely in the regulatory region of a virulence gene (CT115/*incD*) believed to play a role in subverting the host cell machinery. Nevertheless, we found no significant differences in the growth rate, plasmid load, and attachment/entry rate, between strains before and after their long-term laboratory propagation. Of note, from the mixture of clones in E/CS88 initial population, an inactivating mutation in the virulence gene CT135 evolved to 100% prevalence, unequivocally indicating that this gene is superfluous for *C. trachomatis* survival *in vitro*. Globally, *C. trachomatis* revealed a slow *in vitro* adaptation that only modestly modifies the *in vivo*-derived genomic evolutionary landscape.

7.2. Introduction

Random mutation events create variation that is purged by selection in the never-ending process of microbial evolutionary adaptation. The specific causes of variation are hardly traceable as they mostly depend on fluctuations in the environment, which may be as diverse as immune pressure, lack of nutrients, or changes in physiological conditions, such as pH or temperature. In the bacteria research field, it is important to understand the association between the replication associated errors, the frequency of selective sweeps that purge those errors, and the ultimate impact of these dynamics on organism adaptation to new environments. For DNA-based microbes, reported mutation rates *per* base pair *per* generation vary up to four orders of magnitude, and high mutation rates have been found in laboratory asexual bacterial populations (419,420).

Several studies have evaluated the microbe genomic alterations due to laboratory passaging but limited data are available so far for obligate intracellular bacteria (421,422), where the interaction with the host cell line governs the evolutionary process. Indeed, according to the Red Queen

Hypothesis, interacting species undergo an arms race of continuous adaptation and counter-adaptation that drives molecular evolution (423). By extrapolating this scenario to the *in vitro* system, this constitutes a rational basis for assuming that bacterial strains “adapted” to the laboratory for a long time shape their genome in a different fashion from the wild-type strains. Indeed, the latter deal with different biological host niches encompassing dissimilar immunological and physiological status, as well as competing microbiota. This is reflected by the bacterium *Chlamydia trachomatis*, a human pathogen that causes non-invasive (restricted to the mucosal epithelium of ocular or genital tract) or invasive (mainly through the dissemination to lymph nodes leading to an inguinal lymphadenopathy named lymphogranuloma venereum - LGV) diseases. This obligate intracellular pathogen has a unique infectious cycle of about 48 hours (424), and alternates between an infectious form (elementary body - EB) and a replicative form (reticulate body - RB) that multiplies within a host vacuole named inclusion. The culture isolation of *C. trachomatis* dates back to the 1950s (192) allowing the isolation of “prototype” strains, which were collected decades ago and have been laboratory maintained since then. These historical isolates have been focused by thousands of studies as they are accessible to laboratories worldwide, allowing cross-comparison and complementation of results. However, the genomic stability of these strains has been questioned as there is the assumption that the laboratory passaging has been affecting their genomic make-up (425).

It was our goal to evaluate the impact of laboratory propagation on *C. trachomatis* genome dynamics. We performed a long-term serial *in vitro* expansion of both a prototype strain with a long history of laboratory passaging and a clinical isolate collected from a woman with pelvic inflammatory disease (that has not been *in vitro* passaged) during over a year (up to 3500 bacterial generations), and compared the full-genome sequences before and after the strains’ propagation.

7.3. Materials and Methods

7.3.1. Long-term C. *trachomatis* propagation in tissue culture

The present study involved the long-term *in vitro* propagation of two *C. trachomatis* strains in the generally used HeLa229 cell line: the historical prototype strain L2/434/Bu (invasive genital strain/serovar L2) and the clinical isolate E/CS88 (non invasive genital strain/serovar E). L2/434/Bu was isolated in 1968 in California (USA) from an inguinal bubo (426). This strain was initially accessioned as an egg passaged preparation, and subsequently maintained in human cell lines at ATCC since then. We opted to use this prototype strain both because this isolate has been one of the most studied *C. trachomatis* strains worldwide, and to test if an evolutionary different scenario would be obtained when compared with a clinical isolate that had never been laboratory propagated before. In our study we used L2/434/Bu strain obtained from the ATCC collection. The strain E/CS88 belongs to the culture collection of the Portuguese National Institute of Health, and was isolated in 1993 from a woman with chlamydial pelvic inflammatory disease. The diagnostic was performed by COBAS Amplicor PCR (Roche Molecular Systems, Branchburg, NJ, USA) and culture, as previously described

(427), and both the original endocervical swab and the harvested infected HeLa229 cells were stored in liquid nitrogen since then.

The initial inoculation of strains was generally performed using previously described techniques (231). Briefly, each strain was inoculated onto confluent *Mycoplasma*-free HeLa229 cell monolayers (in T25 cm^2 flasks) by centrifuging for 1 h at 34°C at 2200 rpm. Subsequently, the cultures were incubated for 1 h at 37 °C, 5% CO_2, the cell medium was replaced by an enriched medium (MEM 10% foetal bovine serum, vitamins, non-essential aminoacids, glucose and 0.5 µg/ml cycloheximide), and cultures were allowed to grow at 37 °C, 5% CO_2. The inclusion development was regularly monitored through phase-contrast microscopy by checking the inclusion size in order to precisely decide when the passage is needed. In a continuous fashion, the bacterial-saturated cells were harvested by trypsinization and transferred to new T25 cm^2 flasks containing a suspension of fresh HeLa229 cells (which enable the still-dividing RBs to proceed with the developmental cycle). When necessary, and specifically in order to avoid the culture decline, the suspension of infected-cells was sonicated (Vibra Cell, Bioblock Scientific) for disrupting eukaryotic cells and bacterial releasing. Subsequently, the cell debris was discarded through low-speed centrifugation, and the *Chlamydia*-enriched supernatant was then inoculated by centrifugation (as described above). As our goal was to study the mutations generated during the replicative process in the laboratory propagation, we adopted trypsinization as the major propagation method because sonication discards dividing bacteria. The regular microscopy visualization allowed us to make sure that the inclusions continued their regular development after trypsinization, although we cannot exclude that some inclusions may be lost. Both L2/434/Bu and E/CS88 cultures were serially maintained for over one year, which constituted about 250 tissue culture passages. Considering the use of different harvesting methods as well as the fluctuations on growth rate observed for both strains during the long-term experiment, the passages were not performed at rigid time intervals (frequently, passages were spaced less than 36 h).

7.3.2. Preparation of DNA for sequencing

For reading simplification and when the need arises, an “(i)” or “(f)” will be added to the designation of the strains. Thus, L2/434/Bu(i) and E/CS88(i) represent the strains in the initial stage of the evolution experiment, whereas L2/434/Bu(f) and E/CS88(f) refers to the final stage. For each *C. trachomatis* strain, we performed a scale-up from both initial and end-point cultures in order to generate sufficient quantities of DNA for full-genome sequencing. *Chlamydia*-enriched cells in up to 10 T75 cm^2 flasks were harvested by using glass beads, re-suspended with cold phosphate-buffered saline (PBS) and subjected to a discontinuous density gradient purification procedure adapted from previous studies (428,429) to obtain chlamydial material free of host-cell contaminants. Briefly, cell suspensions were pooled, ruptured by sonication and centrifuged at 500g for 10 min at 4 ºC. The resulting supernatant was further centrifuged at 30000g for 1h at 4 ºC. The final pellet was re-suspended in PBS, homogenized by sonication, and finally pipetted over layers of discontinuous urographin [Urographin 76%: sodium amidotrizoate (0.1 g) and meglumine amidotrizoate (0.66 g);

Bayer - Portugal)] gradients: 3 ml of 60%, 10 ml of 52%, 10 ml of 45% and 10 ml of 30% urographin (vol/vol) in 40 ml tubes. These gradients were centrifuged at 26500g for 1.5h. The EB fraction (located at the 45/52% urographin interface) and the RB fraction (located at the 30/45% urographin interface) were collected, diluted in PBS, and then centrifuged at 30000g for 1 h at 4 ºC. The resulting pellets were washed with PBS to remove residual urographin, and re-suspended in PBS. DNA was extracted using the DNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. Prior to sequencing, several assays were performed to assess the purity and quality of DNA recovered by this method. We preliminary checked the degree of human DNA contamination in both EB and RB fractions by quantifying the copy number of *C. trachomatis per* copy number of HeLa229 genomes using a previously described real-time quantitative PCR (273). We opted to use exclusively the EB fractions as we found a degree of contamination with host-cell DNA of at least 70-fold lower in the fraction of EBs (less than 0.01% of contaminant host-cell DNA copies in each EB sample). The quality of DNA was verified by agarose gel electrophoresis, and A260/A280 readings yielded optimal results of ~1.8. Finally, the DNA was quantified by fluorometry (Quant-iT Picogreen, Invitrogen), and ~500 ng of highly pure DNA from L2/434/Bu(i), L2/434/Bu(f), E/CS88(i) and E/CS88(f) were used for full-genome sequencing.

7.3.3. DNA sequencing and assembly

The chromosome and plasmid of both L2/434/Bu and E/CS88 were sequenced before and after their intensive continuous laboratory passaging by using a Roche 454 GS FLX Titanium according to the standard manufacturer's instructions (Roche-454 Life Sciences, Branford, CT, USA) at Biocant (Cantanhede, Portugal). A mean of 163591 reads (after removing vestigial reads resulting from contamination with human and *Mycoplasma* DNA) with an average length of 443 bp were generated for the four samples, representing a coverage between 56- and 75-fold (Supplemental Table S7.1). The reads were mapped through the GS Reference Mapper v2.6 software, using as reference the available sequences of previously full-sequenced *C. trachomatis* strains. For the analysis of L2/434/Bu, as this strain had already been sequenced in 2008 (46), the available genome and plasmid sequences (GenBank accession number NC_010287 and NC_010285, respectively) were used. For the E/CS88 isolate, we used the available chromosome sequence from the strain E/Bour (GenBank accession number HE601870.1) and the plasmid sequence of the strain E/SotonE4 (GenBank accession number HE603232.1). Given the non-repetitive nature of the *C. trachomatis* genome, a low number of contigs (no more than five) were obtained for each sample, where the plasmid sequence was presented as a single contig. Globally, the contigs covered more than 99.4% and 100% of to the length of the chromosome and plasmid sequences of the references, respectively (Supplemental Table S7.1). All variant positions between the samples and the references revealed by the GS Mapper software were carefully inspected. Only unambiguous base variants [i.e., noticed by both forward and reverse readings in a similar percentage of reads and not laying in the 454/Roche error-prone long homopolymeric regions (46)] were considered. Finally, the chromosomal consensus contigs were simply joined taking into account the sequence of the contig boundaries and the genome

of the reference sequences. The sequences were submitted to GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/index.html), and are available under accession numbers CP003963-CP003966 (for L2/434/Bu) and CP003970-CP003971 (for E/CS88 plasmid). For E/CS88 chromosome, whole genome shotgun projects have been deposited under the accession ANOM00000000-ANON00000000. The versions described in this paper are the first versions, ANOM01000000-ANON01000000.

7.3.4. Identification of mutations due to laboratory passaging

The mutations detected between sequences retrieved from the initial and final stages were classified according to their frequency in the bacterial population. Thus, when base variants were supported by all sequence reads, they were considered as fixed mutations (i.e., carried by all clones in the population). On the other hand, as the high achieved coverage (56- to 75-fold) allowed the identification with great confidence of nucleotide positions revealing mixture of clones, the corresponding mutations were classified as: *i*) high-frequent (more than 50% of reads support the mutation, i.e., carried by the majority of the clones); or *ii*) low-frequent (less than 50% of the reads carry this nucleotide change, i.e., carried by low-frequent clones that have risen to a measurable frequency). For each identified mutation, we further assessed its chromosome location, structural class (transversions or transitions) and its effect at amino acid level (synonymous or non-synonymous mutations) based on fully-sequenced *C. trachomatis* strains for which complete annotation was performed (46,115,165).

7.3.5. Impact of laboratory propagation on *C. trachomatis* fitness

We evaluated differences in *C. trachomatis* fitness before and after laboratory propagation by evaluating the growth rate, plasmid load *per* strain and attachment/entry rate. All these assays were performed in duplicate. The growth kinetic of the strains under evaluation was assessed through the calculation of the bacterial generation time (doubling time), which also allowed the extrapolation of the total number of bacterial generations reached during the ~1 year experiment. For this, we collected samples from four stages of the experiment including the initial and the final stages. The inoculations were preceded by sonication as a synchronous cycle was required for this specific assay, in order to guarantee that the whole bacterial population was in exponential growth at the time-points chosen for DNA extraction (10, 20, 30 hours post-infection). DNA samples were used to quantify the number of *C. trachomatis* genomes throughout real-time quantitative PCR using the ABI 7000 SDS (Applied Biosystems) as performed in a previous work (231). As for other bacteria, the doubling time (dt) was calculated using the standard formula: $dt = (t_f - t_0)/n$, where $t_f - t_0$ is the time elapsed during the period of higher growth and n is the number of bacterial generations. The n was calculated based on the formula: $n = (\log_{10}N_f/N_0)/\log_{10}2$, where N_f and N_0 are the genome copy numbers determined at the defined time-points.

To quantify the number of *C. trachomatis* plasmids, we used the same DNA samples (10, 20 and 30 h post-infection) as above, and performed absolute real-time quantification. Briefly, a 51bp fragment of a single-copy *C. trachomatis* plasmid gene (pCTA_0008; according to the plasmid annotation of the strain A/Har13; GenBank accession number CP000052) was amplified (primers pCTA_0008_1 ATTTTTCCGGAGCGAGTTACG and pCTA_0008_2 GTACATCGGTCAACGAAGAGGTT designed using Primer Express software - Applied Biosystems) and cloned into a TOPO vector before transforming DH5α competent cells. High-copy-number plasmid DNA was purified from selected transformed-colonies, and the exact number of plasmids was measured. Cloning, transformation, and plasmid purification and quantification procedures were performed as previously described (231). Finally, the obtained plasmid copy number was divided by the corresponding number of chlamydial genomes in each DNA sample.

For the evaluation of the attachment/entry rate, we used fresh EB-enriched inocula (twin-inoculums were stored at -80 ºC) to infect confluent HeLa229 cell monolayers in 6-well culture plates (multiplicity of infection of 1). After centrifuging for 1 h at 3500 rpm, plates were placed at 37 °C (5% CO_2) for 1 h. Subsequently, the inoculum was removed, the cells were washed with fresh medium (to remove EBs that were not capable to infect), and plates were again placed at 37 °C (5% CO_2) for 1 h. Cells were scraped and mechanically disrupted by sonication. DNA was extracted from both the initial inocula and the collected samples at 2 h post-infection (to ensure that no “second-generation” bacteria would be quantified) using the DNA Mini Kit (Qiagen) according to manufacturer's instructions. Finally, the number of chlamydial genomes was quantified by real-time PCR in each sample, as described above. The attachment/entry rate was then evaluated for L2/434/Bu(i), E/CS88(i), L2/434/Bu(f) and E/CS88(f) by measuring the fraction of bacteria present in the initial inoculum that effectively infected HeLa229 cells.

7.3.6. Mathematical modeling

Considering the mutational profile revealed in the present long-term evolution experiment, we performed mathematical modeling in order to study the selective advantages underlying variations in the frequency of mutations in evolved bacterial populations. We calculated the selective advantage (*S*) for a specific mutation occurring in the clone *b* (derived from the non-mutated clone *a*) as:

$$S = [(Pa_0/Pb_0) \times (Pb_n/Pa_n)]^{1/n}$$

where *Pa* and *Pb* are the percentages of the non-mutated and the mutated clone in the population (*Pa* + *Pb* =1), respectively, in the interval of bacterial generations (0, *n*). By modeling the fixation of a single mutation, the putative existence of clonal interference (i.e., competition between clones carrying dissimilar advantageous mutations) is not taken into account, which could somehow increment the selective advantage values displayed by this model. Also, other variables such as mutation rates *per* type of mutation, which are well-studied for other microbes (430), were not included because the association between mutant genotypes and phenotype is not straightforward in *C. trachomatis*. The present simulation also allowed us to infer if the frequency of different clones present in the initial population (which was used for full-genome sequencing) was adulterated during the short scale-up

procedure from the clinical swab. Furthermore, based on the calculated mutation rates (see results), we predicted the rate of appearance of variant clones during a single *C. trachomatis* developmental cycle.

7.4. Results

7.4.1. Clonal characterization of the initial populations

We characterized the clonal diversity present in the bacterial population of the starting inocula as no *a priori* clone isolation was performed. For the prototype strain L2/434/Bu, the present evolution experiment likely constituted an expansion of a single clone as no measurable mixture of clones was detected in any genomic position. As a curiosity, we compared the obtained genome sequence with that from a same-ancestor clone fully-sequenced in 2008 (46), and detected only two chromosomal nucleotide differences: *i*) the T↔C in the intergenic region IGR CTL0031(*hemA*)-CTL0032(*sycE*) (chromosome position: 39654 bp); and *ii*) the non-synonymous mutation C↔T in the gene CTL0259 (*gatB*) (chromosome position: 326941 bp).

Contrarily, for the clinical isolate E/CS88, the population in the initial inoculum was found to be a mixture of clones. The full-genome data exposed 23 nucleotide positions revealing clone variants in the population (Supplemental Table S7.2). A detailed analysis revealed a bacterial population composed by one clearly predominant clone (represented in more than 70% reads) and a non-measurable number of low-frequent variant clones. It is worth noting that from the 23 variable nucleotide positions, two of them affect the same gene (CT530/*fmt*), and three are located in genes (CT135, CT394/*hrc* and CT511/*rlpO*) where mutations had been previously found among clones from a same parental serovar D prototype strain (202). In particular, the gene CT135 has been pointed out as a mutagenic hot-spot for *Chlamydia*e (431), since multiple CT135 genotypes were found among mixed-clone populations in both *C. trachomatis* (202,432) and *Chlamydia muridarum* (32,431) species. Also, most of the reported polymorphisms in CT135 were single-nucleotide indel events or single inactivating mutations that prematurely truncate the predicted proteins. Remarkably, the predicted CT135 protein of the dominant clone of the strain E/CS88(i) is also truncated by a premature stop codon leading to a loss of 257 amino acids from its C terminus, which likely causes the complete loss of the protein function.

7.4.2. *C. trachomatis* growth rate

The bacterial doubling time of the strains under evaluation was estimated in order: *i*) to extrapolate the number of bacterial generations enrolled during this long-term experiment; *ii*) to calculate the mutation rates *per* base pair (or *per* genome) *per* replication; and *iii*) to evaluate the bacterial fitness throughout the assessment of growth dynamics in the initial and final stages of the long-term *in vitro* propagation. Throughout the experiment, we estimated mean doubling times of 2.78 h (SD±0.56) and 2.72 h (SD±1.41) for the L2/434/Bu and E/CS88 strains, respectively, which are in the range interval (1.5 to 4 h) of previously reported doubling times for this species (98,231,433,434).

Although LGV strains are generally assumed to grow faster than remainder strains, several reports have found either non-LGV strains growing as fast as the former or discrepancies in the calculated doubling time for the same LGV strain (231,424,434). The large standard deviations reflect the fluctuations in the growth rates observed during the ~1-year experiment, and also bias the extrapolation of the number of generations elapsed, ranging from ~2000 to ~5000 with a mean value of 3500 generations. Although there were permanently bacterial populations under exponential growth in the continuously maintained cultures (due to the use of trypsinization as the major harvesting method), these estimates are somehow biased by the existence of non-replicating sub-populations. Nevertheless, we believed this is a rational approach considering the asynchronous *C. trachomatis* life-cycle (98).

7.4.3. Evaluation of genomic alterations after long-term *in vitro* passaging

The whole-genome sequence data retrieved from the *C. trachomatis* cultures at end of the long-term passaging experiment were compared with those obtained from the initial inocula. Regarding the plasmid, no changes were observed in the sequences for both L2/434/Bu and E/CS88 strains. For the chromosome, neither indel or duplication events nor genomic rearrangements were observed, but we detected few single nucleotide mutations (SNPs) that were introduced in the bacterial population of both strains during laboratory propagation (Table 7.1). The mutations were classified according to their frequency in the population at the time of sampling: *i*) fixed (present in all clones of the population); *ii*) high-frequent (carried by most clones); and *iii*) low-frequent (mutations at low measurable frequency).

For the prototype L2/434/Bu, we have initially propagated a single clone but the resequencing at the end of the experiment revealed a mixture of coexisting clones in the population. In fact, besides two mutations that swept to fixation (affecting the *loci* CTL0084 and CTL0103), we observed four additional chromosomal nucleotide positions variable among clones: two mutations [in the IGR tRNA-Leu/CTL0243(xerD) and CTL0486] also carried by the high-frequent clone, and two mutations (in CTL0610 and CTL0818) from clone variants at a low measurable frequency in the population. The mutations are essentially in coding regions, alter the coded amino acid, and A.T ↔ C.G transversions are overrepresented (Table 7.1). For the E/CS88 strain, for which we had identified one clearly predominant clone in E/CS88(i) population, we observed one single clone in the E/CS88(f) population harboring one single fixed mutation [an A.T ↔ C.G transversion affecting the IGR CT114/CT115 (*incD*)] relative to the high-frequent initial clone (Table 7.1). In fact, all high frequent bases of the 23 variable positions initially identified for the E/CS88(i) population (Supplemental Table S7.2) became fixed. This scenario is likely explained by both the extinction of the initially present low frequent clones, and the fixation of one adaptive mutation in the high-frequent initial clone during laboratory passaging. As stated above, a *Mycoplasma* contamination was detected but, due to its residual nature, it is unlikely that it constituted a competitive factor and influenced the observed evolutionary scenario. All nucleotide changes here detected resulting from the long-term laboratory propagation had never been reported in any *C. trachomatis* strain for which full-genome sequences were already made available.

Table 7.1. Genomic alterations after long-term *in vitro* passaging.

Strain	Mutation location [a]	*loci* [b]	nt (aa) change [c]	Ts/Tv [d]	Frequency [e]	Putative function
L2/434/Bu						
	107093	CTL0084 (CT715)	G → C (Arg→Pro)	Tv	F	UDP-N-acetylglucosamine pyrophosphorylase(115).
	127347	CTL0103 (CT734)	A → G (silent)	Ts	F	Lipoprotein (46).
	297976	IGR tRNA-Leu/CTL0243 (*xerD*) [IGR tRNA-Leu/CT864 (*xerD*)]	C → A	Tv	H	----
	579560	CTL0486 (CT234)	A → C (Gln→His)	Tv	H	Membrane transport protein from the Major Facilitator Superfamily (46,115).
	727251	CTL0610 [CT356 (*yyaL*)]	A → C (Asn→His)	Tv	L	Thioredoxin domain-containing protein (115,435)
	946412	CTL0818 (CT555)	G → T (Ala→Ser)	Tv	L	snf2/rad54 family helicase C-terminus (435)
E/CS88						
	135236	IGR CT114/CT115 (*incD)*	A → C	Tv	F	----

[a] Polymorphism location refers to the location in the reference genomes: L2/434/Bu (GenBank accession number NC_010287) and E/Bour (GenBank accession number HE601870.1).

[b] The *loci* designations are based on genome annotation of the strain L2/434/Bu (GenBank accession number NC_010287). The corresponding nomenclature in the annotated D/UW3 genome (GenBank accession number NC_000117) is also shown between parentheses. For E/CS88, only the latter was used. IGR, intergenic region with adjacent ORFs indicated.

[c] The nucleotide changes in open reading frames are presented in the 5' to 3' direction.

[d] Ts and Tv correspond to A.T ↔ G.C transitions and A.T ↔ C.G transversions, respectively.

[e] Three levels of the frequency of mutations in the clone-mixture population were established: i) fixed (F): mutations carried by all clones (all reads support the mutation); ii) high-frequent (H): mutations carried by the majority of the clones (more than 50% of reads support the mutation); and iii) low-frequent (L): mutations at low measurable frequency (less than 50% of the reads carry this nucleotide change).

7.4.4. Impact of laboratory propagation on *C. trachomatis* fitness

While monitoring both chlamydial cultures by phase-contrast microscopy throughout the long-term experiment we observed no visible changes in the size of inclusions as well as in cell morphological features previously pointed to be indicative of differences of strains' cytotoxity, namely cell rounding, detachment, and lysis (436). We also evaluated the bacterial doubling time as it is expected to be a stable and species-specific feature under optimal growth conditions, where fluctuations in the growth rates may be indicative of fitness alterations (Andersson and Hughes, 1996). Although we observed visible fluctuations on the strains growth rate (demonstrated by variable

doubling times calculated during the experiment, as described above), L2/434/Bu(f) and E/CS88(f) presented growth rates resembling those of L2/434/Bu(i) and E/CS88(i), respectively.

We also checked for variations in the number of plasmids *per* genome (Figure 7.1) as the plasmid was recently pointed to play a primary role in *C. trachomatis* infectivity *in vivo* (379). For both strains, higher plasmid copy numbers were observed during the exponential phase of the life-cycle, and no significant differences were detected in the number of plasmids *per* genome between the initial and final cultures (P=0.268 and P=0.075 for L2/434/Bu and E/CS88, respectively; two-way ANOVA test), indicating that the plasmid load was likely not affected by the long-term *in vitro* maintenance of *C. trachomatis* strains.

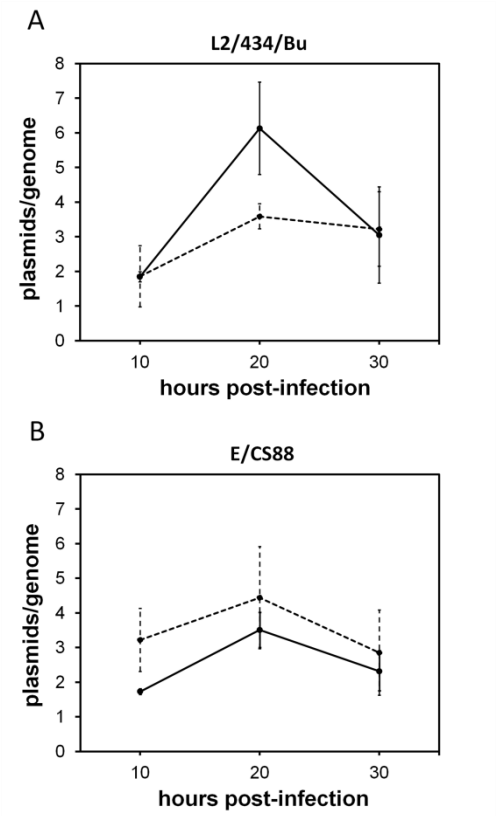


Figure 7.1. Impact of long-term laboratory passaging on the *C. trachomatis* plasmid load. The number of plasmids *per* genome was evaluated before and after the laboratory propagation of the strains L2/434/Bu (A) and E/CS88 (B). The plasmid load was quantified throughout the developmental cycle (10, 20 and 30 hours post-infection) of the strains by absolute real-time quantitative PCR. Solid and dashed lines represent the strains in the initial and final stages of the evolution experiment, respectively. The results (mean ± SD) are based on two independent experiments.

Finally, as measure of the attachment/entry rate, we have also compared the percentage of chlamydial organisms that effectively infected HeLa229 cells using bacterial inocula from both the initial and final cultures (Figure 7.2). For the historical strain L2/434/Bu, we detected similar levels of attachment/entry rate for L2/434/Bu(i) and L2/434/Bu(f) (*P*=0.868; t-Student test), whereas E/CS88(f) displayed an increase of attachment/entry rate of about 20% when compared to E/CS88(i) (*P*=0.057; t-Student test).

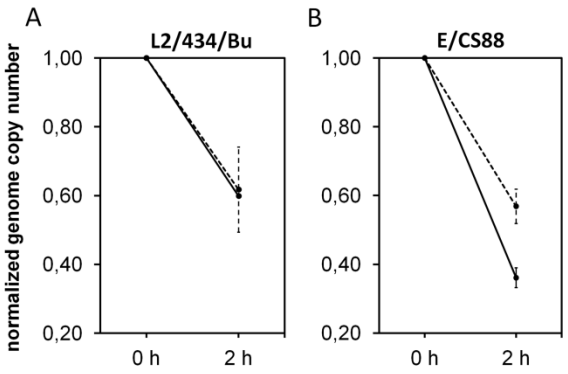


Figure 7.2. Evaluation of the attachment/entry rate of strains before and after their long-term *in vitro* propagation. The attachment/entry rate was evaluated by calculating the fraction of bacteria present in the initial inoculum (measured at 0h) that effectively infected HeLa229 cells (measured at 2h post-infection). The assay was performed before (solid line) and after (dashed line) the long-term laboratory propagation of the strains L2/434/Bu (panel A) and E/CS88 (panel B). The number of *C. trachomatis* genomes in each sample was quantified through absolute real-time quantitative PCR. The values (mean ± SD) at 2h post-infection were normalized to the number of genomes present in the initial inoculum (0 h), which were arbitrarily set to 1. The ~20% increase of attachment/entry rate observed for E/CS88(f) when compared to E/CS88(i) was not statistically significant (*P*=0.057; t-Student test). The results are based on two independent experiments.

7.4.5. Calculation of mutation rate and mathematical modeling

The *C. trachomatis* mutation rate *per* base pair *per* replication (μ_{bp}) were 3.84×10^{-10} for E/CS88 and 1.10×10^{-9} for L2/434/Bu, which correspond to a mutation rate *per* genome *per* replication (μ_g) of 0.0004 and 0.0011, respectively (i.e, 0.04-0.11% of the new clones generated after one single replication in a bacterial population will carry a new mutation). By extrapolating these mutational dynamics to a single developmental cycle (Figure 7.3), it could be virtually expected that up to 0.25% of the generated clones will be variant. It is worth noting that this value is underestimated as there is no way to take into account mutations that do not reach a detectable frequency (including deleterious mutations, which rapidly disappear) (437).

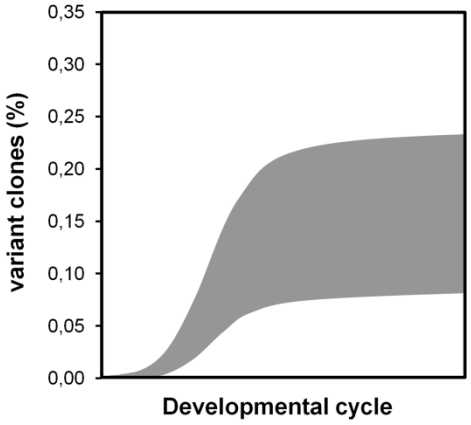


Figure 7.3. Prediction of the rate of appearance of variant clones. The graph shows the percentage of the variant clones that likely arise during one single *C. trachomatis* developmental cycle (about seven bacterial generations). Values were extrapolated from the empirically estimated mutation rate *per* genome *per* replication (between 0.0004 and 0.0011) and are independent of the initial population size.

We also performed mathematical modeling to numerically plot the selective advantage of one mutated clone against the number of generations needed to reach a given frequency in the population (Figure 7.4A). In fact, for an asexual evolving bacterial population under strict neutrality, it is intuitively expected that one clone carrying a new mutation will not rise in frequency unless the mutation confers advantage over its ancestor and other lineages. For a specific final frequency, the selective advantage is a decreasing function of the number of generations. Also, for a fixed number of generations, higher final frequencies implicate higher selective advantages of the mutated clone. We further applied this model to assess if the clone frequencies in the E/CS88(i) population (revealing a dominant clone with >70% frequency) reflect those contained in the clinical swab. In fact, we cannot discard that deviations in these frequencies could have been introduced during the minimal culture scale-up (four *in vitro* passages) performed to obtain sufficient DNA for full-sequencing of the initial inoculum. The computational simulations presented in the Figure 7.4B revealed that if the dominant clone was a minor frequent clone in the population of the original swab, by assuming for instance 5%, it would need an overwhelming selective advantage of more than 1.14 to reach (after solely four *in vitro* passages) the empirically observed 70% frequency at the time of sequencing (Figure 7.4B). This value seems somehow unrealistic considering the model highlighted in Figure 7.4A. Also, this would reflect an opposite scenario from the one observed in the present long-term experiment. In fact, although up to 0.25% of variant clones are expected after each *C. trachomatis* developmental cycle (Figure 7.3), only a single advantageous clone was observed for the E/CS88(f) population after thousands of generations, which sustain the well-assumed scenario where a random mutation have a considerable lower probability of being advantageous than neutral or deleterious in evolving bacterial populations (419). In an alternative hypothesis, a dominant nature-rescued clone (hypothetically presented in the bacterial population in the clinical swab) would have undergone a complete extinction throughout the

minimal scale-up process due to deleterious effects of the *in vitro* environment. However, this is also hardly explained considering the likely optimal physiological conditions underlying cell culture. Accordingly, we believe that the clone frequencies in the E/CS88(i) population reflect those contained in the clinical swab. Nevertheless, the only way to really prove this would be the immediate use of the clinical sample for sequencing (which is unfeasible due to insufficient number of bacteria in the swab).

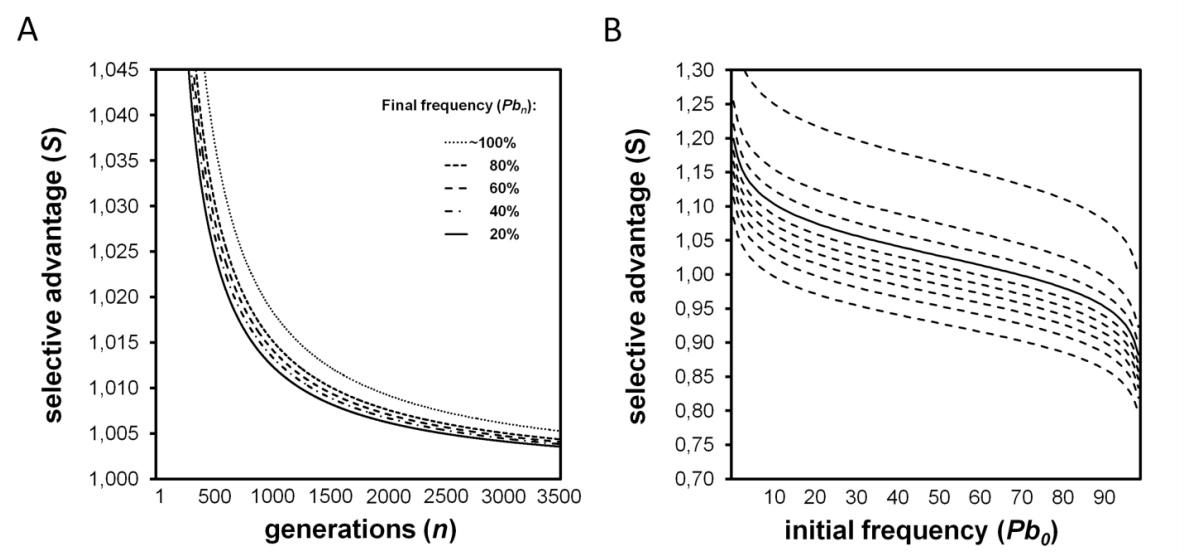


Figure 7.4. Mathematical modeling of fluctuations in clone frequency. A. The number of generations (*n*) needed by a mutated clone to reach a given final frequency (Pb_n) is governed by the selective advantage (*S*) conferred by a beneficial mutation. Higher selective advantages implicate lower number of generations needed, or higher frequencies reached (for a fixed number of generations elapsed). This simulation assumes a population size of 10^6 at the time the advantageous mutation occurred. B. This simulation reflects the selective advantages underlying putative fluctuations in the initial clones' frequency (Pb_0) during the minimal culture scale-up procedure (four *in vitro* passages) (*n*=30). The solid line corresponds to the empirically observed frequency (70%) of the abundant clone in the E/CS88(i) population, whereas the dashed lines represent final frequencies (Pb_{30}) from 10% (lower plot) to ~100% (upper plot).

7.5. Discussion

It is well known that pathogens may adapt to laboratory culture conditions, which implies that sequences obtained from laboratory-maintained strains may contain artifacts that introduce bias when analyzing evolutionary landscapes. This has been observed, for instance, in human influenza virus, where "laboratory-derived" mutations (mainly in the receptor binding pocket of the haemagglutinin protein) were found to account for up to 8% of all non-silent changes (438), as well as in the gp120 envelope glycoprotein of HIV-1 (439), and in the VP1 capsid protein of foot-and-mouth disease virus (440). In *C. trachomatis*, the fixation of mutations in the *in vitro* environment was experimentally demonstrated through the induction of selective pressure by antibiotic exposure (441-445). Previous studies that attempted to evaluate if *in vitro* propagated *C. trachomatis* strains are genetically unstable focused on analyzing the stability of the traditional typing gene - *ompA* (over 20 laboratory passages)

(422) and of loci from multi locus typing systems (over 72 laboratory passages) (421), and found no variation on sequence data retrieved before and after the isolates being expanded in tissue culture. Those studies specifically aimed to evaluate the stability of genotyping markers, so the limited number of genomic regions analyzed (up to nine) and the short-term duration of the experiments prevent the knowledge of the real extent of genomic alterations that may arise due to laboratory adaptation.

In the present study, we analyzed the impact of long-term *in vitro* propagation on the genome of two *C. trachomatis* strains by fully-sequencing their genomes before and after about 250 *in vitro* passages in tissue culture (up to 3500 bacterial generations). We observed four SNPs (six if one considers additional two carried by low frequent variant clones) for the historical prototype strain L2/434/Bu and one SNP for the clinical isolate E/CS88 (Table 7.1), which corresponds to an estimated mutation rate in the range of 3.84×10^{-10} - 1.10×10^{-9} mutations *per* base pair *per* generation, a value that fits previously estimates for other bacteria (153,446). Due to the obligate intracellular life-style of *C. trachomatis*, which is characterized by small replicating populations frequently subjected to transmission bottlenecks, we would expect to observe some genetic drift. In fact, we had previously detected traces of Muller's ratchet mutations (150,151) as a result of *C. trachomatis* evolutionary speciation (301). Despite the present evolution experiment enrolled an unprecedented evaluation of thousands of chlamydial generations, this "evolutionary scale" still may be too modest to allow the observation of such evolutionary phenomena *in vitro*. For instance, a similar study in *E. coli* detected a drift signature solely after 20000 generations (447). Considering that this free-living bacterium replicates about eight-times faster than *C. trachomatis,* such evaluation would be unworkable for this bacterium whose growth is host-cell dependent.

We believe that the present experiment encompasses a number of bacterial replications that may represent decades of cell culture work enrolling a specific strain of a common chlamydial research laboratory. Indeed, cultures are commonly performed from stocks of multiple twin-aliquots of enriched inocula, which are used over several years without being further re-inoculated for generating additional aliquots (and thus, without introducing substantial bacterial generations in the initial population). Therefore, our results pointing to a considerable genomic stability *in vitro* suggest that the *in vivo*-derived genetic diversification is not significantly compromised by the "modest" number of laboratory passages performed throughout decades in most chlamydial research laboratories. Accordingly, when we compared the full-genome sequence of our L2/434/Bu(i) with that of L2/434/Bu from another research laboratory, which was fully-sequenced five years ago (46), we only found two nucleotide differences, although these clones enclose necessarily distinct histories of laboratory maintenance.

We performed computer simulations to study mutational fitness, which suggest that alterations in frequencies of the clone variants in the population are possible only under non-neutral selective pressures. Indeed, in a random mutational scenario exclusively mathematically ruled (i.e., in the absence of evolutionary forces driving selective sweeps), any frequent mutation in the beginning of the experiment will keep its frequency at the end. Thus, we speculated that the mutations observed in the present study are advantageous, which is clearly supported by both the mathematical modeling (Figure 7.4) and also the huge amount of literature data concerning the evaluation of laboratory

adaptation of microbes, as comprehensively compiled in a recent review (448). Although phenotype variations are hard to trace due to the lack of straightforward methods for genetically manipulating *C. trachomatis* chromosome, a careful inspection of the observed mutations (Table 7.1) points in that direction. For L2/434/Bu: *i*) all but one mutation are A.T ↔ C.G transversions, which we estimated to be about 2.6 fold more unlikely than A.T ↔ G.C transitions in the *C. trachomatis* diversification [based on previously published data (301)]; *ii*) all but one mutation are located in coding regions and four of them cause amino acid replacement; and *iii*) one of the fixed mutations (non-synonymous) affects CTL0084 that codes for a protein (UDP-N-acetylglucosamine pyrophosphorylase) with transferase activity, which was found to be among the gene functional categories more prone to undergone advantageous mutations in laboratory evolution studies in *E. coli* (420); *iv*) another non-synonymous mutation involves a thioredoxin coding gene (CTL0610); thioredoxins are part of the gene expression regulation apparatus (449), which is known to be a preferred selection target of adaptive mutations (448,450). For E/CS88, the final population had no diversity indicating that the mutation likely conferred a selective advantage, and swept to fixation during the evolution experiment (Figure 7.4). Moreover, the mutation is a A.T ↔ C.G transversion, and it is located in the intergenic region upstream of the gene encoding the inclusion membrane protein D (CT115/*incD*), which is an effector protein (132) recently suggested to play a role in subverting the non-vesicular lipid transport machinery of the host cell (293). However, besides its putative location in the regulatory region of a virulence gene, there is no additional experimental evidence suggesting that the mutation is advantageous. Finally, for both strains, the mutations that became prevalent had to “survive” the drift imposed by the consecutive bottlenecks created during serial passaging, and also the putative occurrence of clonal interference (430). Nevertheless, we cannot discard that some of these mutations can be just neutral mutations that rose in frequency either by their occurrence in the same clone carrying a beneficial mutation or by recombination (genetic hitchhiking) (451). We cannot be conclusive about this because the evaluation of the frequency of selective sweeps would demand sequencing the strains at intermediate stages of the long-term experiment.

In contrast to L2/434/Bu, the initial inoculum of E/CS88 was constituted by a mixture of clones, where the dominant clone (which additionally acquired the described mutation) became 100% frequent after laboratory propagation, and the remainder evolved to extinction. Remarkably, only the latter carried the virulence gene CT135 in its active form. Thus, we concluded that CT135 is superfluous for *C. trachomatis* survival *in vitro.* This strongly corroborates the findings of Sturdevant and colleagues (202) that implicated CT135 in the pathogenesis of *C. trachomatis* urogenital infections while suggesting its dispensability *in vitro*.

Globally, we point to a slow adaptation of *C. trachomatis* to the *in vitro* environment, which, in the present study, was governed by minor genomic alterations. Despite the low number of observed mutations, we cannot discard that they may be sufficient for altering fitness. In fact, marked changes in phenotype were shown to be mediated by very few mutations in studies using *E. coli* (437,452), *Saccharomyces cerevisiae* (453), and the influenza virus (454). Commonly, the phenotype-genotype relationships of mutations are studied through genetic knockout assays, which are still giving the first steps in *Chlamydia* (206,207,213,216,217). Consequently, we performed humble approaches for

evaluating the impact of laboratory propagation on *C. trachomatis* fitness, and the results were not helpful. Indeed, despite fluctuations in the growth rates during the experiment, we did not observe considerable differences between initial and long-term propagated strains. Also, we found no significant variation in the number of plasmids *per* bacterium after the intensive serial passaging of both strains in tissue culture (Figure 7.1), which is intriguing in the light of previous studies suggesting that the plasmid is not required for *C. trachomatis* growth *in vitro* (208,209,379). However, additional studies in *Chlamydiaceae* suggested that the plasmid is important for the infectivity *in vitro* (431,455), indicating that the basis for plasmid maintenance by this species still remains to be clarified. We speculate that an even longer evolutionary experiment would be needed to evaluate the putative *in vitro* extinction of the *C. trachomatis* plasmid, although a recent study in *Borrelia burgdorferi* sensu lato (456) showed the plasmid loss after just less than 10 *in vitro* passages. Regarding the assay for evaluating the attachment/entry rate (Figure 7.2), although we observed an increase in the fraction of bacteria that effectively infected the HeLa229 cells after the E/CS88 strain has been long-term propagated, the result was not statistically significant (P = 0.057; t-Student test). We have no reasonable explanation for this difference, besides the fact that the E/CS88(i) population constituted a mixture of competing clones (with expected variable fitness), and that the E/CS88(f) harbored a putative adaptive mutation.

Although the present assessment is genome-based, it is worth noting that gene expression changes may also account for adaptation of evolving bacteria (420). Besides the discreet genomic changes that were observed in the present study, we cannot rule out the existence of significant transcriptomic alterations on the course of *C. trachomatis* laboratory propagation. In fact, previous laboratory evolution experiments with other bacteria showed that different endpoints of evolution have distinct underlying gene expression states (450,457-459).

Finally, the demonstration of a remarkable degree of parallel evolution between *Pseudomonas fluorescens* and its phage Phi2 (460) makes us wonder if the scenario observed in our study would be mirrored in the absence of such strict interplay between the host cell and the parasite. In fact, for *Coxiella burnetii* [also an obligate intracellular bacterium for which it was already developed a culture medium that allows growth in the absence of host cells - axenic growth (461)], an interesting work found that virulent strains displayed attenuated growth, less ability to induce splenomegaly, and less elicitation of antibody responses when infecting mice after long-term axenic passaging (462). On behalf of the recent advances in the development of an axenic system for *C. trachomatis* culture (94), it would be interesting to evaluate if the host-cell free *in vitro* propagation of *C. trachomatis* strains somehow shapes their genomic or transcriptomic signatures.

As concluding remarks, although phenotype inferences are hard to trace, our genome-based approach highlighting very few nucleotide changes suggests that *in vitro* propagation of *C. trachomatis* strains acts as an “evolutionary freezer” of the *in vivo* genomic diversity rather than an effective modifier of the evolutionary landscape.

Acknowledgements

This work was supported by the grants PTDC/SAU-MII/099623/2008 from Fundação para a Ciência e a Tecnologia (FCT), and ERA-PTG/0004/2010 from FCT in the frame of ERA-NET PathoGenoMics (to JPG). VB and RF were recipients of Ph.D. fellowships (SFRH/BD/68527/2010 and SFRH/BD/68532/2010, respectively) from FCT. AN was a recipient of a post-doctoral fellowship (SFRH/BPD/75295/2010) from FCT.

Chlamydia trachomatis in vivo to *in vitro* transition reveals mechanisms of phase variation and down-regulation of virulence factors

The contents of this chapter were submitted to a peer reviewed international journal.

Borges V., Pinheiro M., Antelo M., Sampaio D.A., Vieira V., Ferreira R., Nunes A., Almeida F., Mota L.J., Borrego M.J., Gomes J.P.

These results were presented as an oral presentation in the Seventh Biennial Meeting of the *Chlamydia* Basic Research Society, New Orleans (Louisiana, USA), March 29 - April 1, 2015.

Personal contribution

VB contributed to the design of the study, performed bioinformatics and most of the experimental work, interpreted data and wrote the manuscript.

Note:

Part of the results were included in the Master Thesis of Minia Antelo Varela, who helped the author of the present Ph.D. thesis in some experimental procedures. The M.Sc. thesis was entitled "Tropism based analysis of *Chlamydia trachomatis* chromosome", on behalf of the Master Degree in Applied Microbiology of the Faculty of Sciences of the University of Lisbon, 2014.

8. *Chlamydia trachomatis in vivo* to *in vitro* transition reveals mechanisms of phase variation and down-regulation of virulence factors

8.1. Abstract

Research on the obligate intracellular bacterium *Chlamydia trachomatis* demands culture in cell-lines, but the adaptive process behind the *in vivo* to *in vitro* transition is not understood. We assessed the genomic and transcriptomic dynamics underlying *C. trachomatis in vitro* adaptation of strains representing the three disease groups (ocular, epithelial-genital and lymphogranuloma venereum) propagated in epithelial cells over multiple passages. We found genetic features potentially underlying phase variation mechanisms mediating the regulation of a lipid A biosynthesis enzyme (CT533/LpxC), and the functionality of the cytotoxin (CT166) through an ON/OFF mechanism. We detected inactivating mutations in CT713/*porB*, a scenario suggesting metabolic adaptation to the available carbon source. CT135 was inactivated in a tropism-specific manner, with CT135-negative clones emerging for all epithelial-genital populations (but not for LGV and ocular populations) and rapidly increasing in frequency (~23% mutants *per* 10 passages). RNA-sequencing analyses revealed that CT135 deletion impacted the expression of multiple virulence factors, namely effectors known to play a role in the *C. trachomatis* host-cell invasion or subversion (e.g., CT456/Tarp, CT694, CT875/TepP and CT868/ChlaDub1). This reflects a scenario of attenuation of *C. trachomatis* virulence *in vitro*, which may take place independently or in a cumulative fashion with the also observed down-regulation of plasmid-related virulence factors. This issue may be relevant on behalf of the recent advances in *Chlamydia* mutagenesis and transformation where culture propagation for selecting mutants/transformants is mandatory. Finally, there was an increase in the growth rate for all strains, reflecting gradual fitness enhancement over time. In general, these data shed light on the adaptive process underlying the *C. trachomatis in vivo* to *in vitro* transition, and indicates that it would be prudent to restrict culture propagation to minimal passages and check the status of the CT135 genotype in order to avoid the selection of CT135-negative mutants, likely originating less virulent strains.

8.2. Introduction

Experimental evolution studies have exploited several organisms' traits to obtain a real-time window on evolutionary processes, to reveal new genetic targets of selection and, in general, to get insight on the genome-scale dynamics of natural selection within microbial populations, constituting a rare opportunity to deeply uncover the correlation (or lack of such) between genetic and phenotypic signatures (451,463,464). In bacteria, common findings include the detection of genes or mutations underlying phenotypic alterations typically observed in laboratory evolving bacterial populations, such as: evolution of new (or loss of) metabolic capabilities, development of antibiotic resistance or sensitivity, or loss of virulence (namely, reduction of the infective capacity) (420,448,451,463,465-

468). For the latter alteration, it has been also observed that: *i*) *in vitro*-cultured microorganisms are in general less virulent than the corresponding wild-types; *ii*) virulence decreases gradually during *in vitro* culture; and *iii*) virulence can eventually be restored by *in vivo* passage (458,462,469-471). These features have consequently been explored by researchers in order to enlighten pathogenic mechanisms underlying the process of infection, and also uncover new virulence genes. Indeed, loss of function of virulence factors may be possibly related to their dispensability in the *in vitro* environment, reinforcing the assumption that these genes may play an essential and specific role *in vivo*.

Few studies have evaluated the impact of laboratory propagation on the population dynamics of obligate intracellular bacteria (201,350,421,422,472), wherein the evolutionary process is ruled by the interaction with the host. In this regard, for instance, the use of laboratory-passaged strains has been frequently questioned when studying *Chlamydia*, including the major human pathogen *C. trachomatis* (32,195,198-201,422), which are organisms displaying a highly specialized capacity of adaptation to the intracellular life-style. Indeed, these obligate intracellular bacteria have a unique biphasic developmental cycle alternating between an infectious form, the elementary body (EB), and a noninfectious replicating form, the reticulate body (RB). Shortly after entering the target eukaryotic cells, EBs differentiate into RBs, which replicate within a membrane-bound parasitophorous vacuole termed the “inclusion”, until the newly formed infectious EBs are released to infect neighboring cells. *C. trachomatis* exploits the host cell throughout all developmental cycle by subverting critical cellular functions (such as, cytokinesis, apoptosis, host cytoskeleton, nutrient transport, membrane trafficking pathways or immune responses) (121,122,125,126) essentially by effector proteins translocated into the host cytosol or localized into the inclusion membrane (Inc proteins) (103,127,128). The majority of the *C. trachomatis* effectors and Incs described so far were found to be transported by using a type III secretion (T3S) system (103,127,129-132,137,138,324). Despite the high degree of genomic conservation in both the 1Mb chromosome and the virulence regulator 7.5 kb plasmid, strains from different *C. trachomatis* serovars typically present dissimilar tissue tropism, disease outcomes and clinical prevalence. In fact, serovars A-C are normally associated with ocular infections leading to blinding trachoma, strains from serovars D-K are the major cause of bacterial sexually transmitted infections (where serovar E is the most clinically prevalent followed by F), and strains from serovars L1-L3 are the causative agents of typical bubonic lymphogranuloma venereum (LGV) or the LGV-associated proctitis (4-6) [the latter is primarily caused by L2b strains (49)].

In a previous approach to study the *in vitro* population dynamics for this bacterium (350), we observed that culture propagation yielded discreet genomic changes in the absence of specific selective pressures. Despite the advance of culture-independent *C. trachomatis* genome sequencing (194-197), our data suggest that the *in vivo*-derived genetic make-up is not strongly compromised during the classical culture procedures for propagating *C. trachomatis*. However, this discreet scenario of genomic alterations may strongly impact gene expression and virulence. Indeed, it was detected, for instance, that clones carrying an inactivated version of the recently identified virulence gene CT135 [which is believed to function in sustaining a more persistent or chronic infections *in vivo* (202,203)], swept to 100% frequency in the population of an epithelial-genital strain (350). Furthermore,

monitoring the impact of introducing *C. trachomatis in vitro* gains special relevance in light of the recent rampant development of transformation and gene knockout experiments in *C. trachomatis,* which demand culture propagation (206,207,213,216,217). Thus, it is crucial to well characterize the *C. trachomatis in vivo* to *in vitro* transition process in order to avoid bias in the interpretation of downstream mutagenesis-derived data and also because virulence of propagated strains may be attenuated. We now performed an unprecedented experimental evolution study over *C. trachomatis* strains from serovars displaying dissimilar clinical prevalence and representing the three disease groups. We evaluated the dynamics by which mutations appear and their frequency in the evolving bacterial populations through whole-population sequencing at various time-points after *C. trachomatis* introduction *in vitro*. We also investigated transcriptome alterations throughout adaptation. Specifically, we aimed to answer the following questions concerning the *in vitro* adaptation of *C. trachomatis*: *i*) is the duration of bacterial developmental cycle affected?; *ii*) is the adaptive dynamics dependent on the tissue-specificity or disease outcomes of the infecting strains?; *iii*) which are the genomic and transcriptomic events underlying adaptation?; and *iv*) is there parallelism in the adaptive outcomes across strains?

8.3. Materials and Methods

8.3.1. *Chlamydia trachomatis* strains

Six *C. trachomatis* strains belonging to the three disease groups (one ocular strain, four epithelial-genital strains and one strain associated with LGV-proctitis) were used in the present study. The *ompA* genotype was verified for all strains, as previously described (39). The ocular strain C/TW-3 was isolated from the human conjunctiva and was obtained from the American Type Culture Collection (ATCC VR-1477). The remaining strains belong to the culture collection of the Portuguese National Institute of Health. The epithelial-genital clinical isolate Ia/CS190/96 was isolated from a male urethra, whereas the epithelial-genital D/CS637/11, F/CS847/08 and E/CS1025/11 were collected from swabs of the cervix of women with no symptoms, HPV-associated warts, and no available clinical data, respectively. Finally, the LGV strain L2b/CS19/08 was isolated from a anorectal swab of a 29-years old HIV(+) man with proctitis and syphilis.

8.3.2. *C. trachomatis in vitro* propagation and preparation for whole-population full-genome sequencing

Each specimen was inoculated in HeLa229 confluent monolayers and allowed to grow at 37 ºC, 5% CO_2 as previously described (473). At the end of the bacterial developmental cycle, the infected cells were harvested using glass beads and disrupted through sonication (Vibra Cell, Bioblock Scientific), and further submitted to low-speed centrifugation (700 rpm for 7min). The bacterial-enriched supernatant was collected and re-inoculated by centrifugation (2200 rpm for 1h) onto new HeLa229 confluent monolayers (first passage). Each chlamydial culture was continuously passaged in

T25 cm^2 flasks. Inoculations were always preceded by sonication followed by low-speed centrifugation, so that the bacterial life-cycle was continuously synchronized. The volume of transferred inocula ranged from 1/10 to 1/2 of the total suspension of the previous passage (depending on the bacterial-enrichment), limiting drift and random selection of clones derived from bottlenecks. Strains D/CS637/11, E/CS1025/11, F/CS847/08 and L2b/CS19/08 were propagated over 30 *in vitro* passages, whereas strains Ia/CS190/96 and C/TW-3 were propagated until the passage 100. The propagation extension of these two strains relied on the need either to clarify the evolutionary scenario that was being observed (for the former) or to obtain a long-term adaptive landscape for an ocular strain (for the latter), mirroring previous data for one LGV and one epithelial-genital strain (350). Passages ~6, 20, 30, 50 and 100 were selected for whole-population full-genome sequencing (see below), where culture scale-up to obtain enough DNA for sequencing was performed two or three passages before each time-point. The contents of these flasks was harvested and further subjected to a discontinuous density gradient purification procedure, before proceeding with DNA extraction and quantification as previously described (350).

8.3.3. Whole-population full-genome sequencing

DNA samples were used to prepare Illumina paired-end libraries according to the manufacturer's instructions (Illumina Inc., San Diego, CA). Libraries were subsequently subjected to cluster generation and paired-end sequencing (2x250 bp) by using the next-generation sequencing platform Illumina MiSeq (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FASTX (http://hannonlab.cshl.edu/fastx_toolkit/) tools were applied to check and improve the quality of the raw sequence data, respectively. Reads were mapped to *C. trachomatis* chromosome and plasmid sequences available at GenBank using both Bowtie2 (version 2.1.0 [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml]) (351). SAMtools/BCFtools (http://samtools.sourceforge.net/) (353) were applied to call single nucleotide polymorphism (SNPs) and insertions/deletions (indels), which were carefully confirmed through visual inspection using the Integrative Genomics Viewer (version 2.3.12; [http://www.broadinstitute.org/igv]) (474). The assembled chromosome and plasmid sequences obtained from the first time-point were used to map the reads generated from subsequent time-points. The frequency of the mutations in the population at the time of sampling were carefully scrutinized through IGV inspection. The mutations reported were supported by a depth of coverage above 150x, a score of quality above 20 for more than 90 % of the bases sustaining each polymorphism and a relative percentage of reads mapping to either of the two DNA strands (strand bias) ranging from 40 to 60 %. We also inferred the plasmid copy number *per* chromosome throughout the different time-points, based on the ratio plasmid/chromosome taken from the respective depth coverage. Chromosome and plasmid sequences obtained at the first time-point were annotated by the NCBI Prokaryotic Genomes Annotation Pipeline 2.3 and deposited in GenBank (see accession numbers in the Supplemental Table S8.1). For the global evaluation of the whole genomic backbone

of the studied *C. trachomatis* genomes, we compared them with the ones from 52 previously sequenced *C. trachomatis* strains (37), as described elsewhere (365).

8.3.4. Impact of *in vitro* passaging on the *C. trachomatis* growth rate

The growth rate (and doubling times) of bacterial populations was measured at the initial and final stages of the evolution experiment in order to assess the impact of *in vitro* propagation on the bacterial fitness (i.e., the ratio of growth rates of evolved population relative to its ancestral). Briefly, for each bacterial population, HeLa229 confluent monolayers on 6-well plates were inoculated (by centrifuging at 3500 rpm for 1h at 34 ºC) at a multiplicity of infection of about 1 with fresh inocula obtained from the cultures (in T25 cm^2 flasks) under passaging. Cells were harvested for DNA extraction at 4, 10, 20 hours post-infection and at the end of the bacterial life-cycle (ensuring that cell lysis had not occurred). DNA samples were then used to quantify the number of *C. trachomatis* genomes at each time-point through real-time quantitative PCR (qPCR), as previously described (350). The growth rate and doubling times (dt) were calculated using the standard formulas: growth rate = (ln N_1 – ln N_0) / $t_{(1-0)}$ and dt = (t_1 – t_0)/n, where (t_1 - t_0) is the time elapsed during the period of higher growth and n is the number of bacterial generations. The n was calculated based on the formula: $n = (\log_{10}N_1/N_0)/\log_{10}2$, where N_1 and N_0 are the genome copy numbers determined at the defined time-points.

8.3.5. Yield of infectious progeny of the serovar D CT135-positive and CT135-negative isolates in cell culture

In order to complement the data from growth rate assays and to obtain an alternative perspective of the impact of the *in vitro* propagation on the chlamydial replicating capacity, we further evaluated the yields of inclusion-forming units (IFUs) (as a measure of progeny) of the serovar D clinical isolate at beginning of the experiment (here designated as “CT135-positive”) and after CT135 loss (here designated as “CT135-negative”). We grew both the D/CS637/11 CT135-positive and CT135-negative strains in HeLa229 cells, and used serial dilution titration to determine the IFU *per* milliliter of the EB-enriched stocks. Briefly, serial dilutions were inoculated as above in HeLa229 monolayers seeded on 96-well plates, the cells were fixed at 22 h post-infection with methanol, stained with an anti–*C. trachomatis* lipopolysaccharide antibody (Pathfinder), and the inclusions were enumerated by immunofluorescence microscopy. Inocula of equivalent concentration (IFU *per* militer) for both the ancestral and the evolved strain were then applied to infect HeLa229 monolayers (on 6-well plates) at a MOI of 1. At the time-points 6, 12, 18, 24, 30, 36, 42 and 48 post-infection, the cultures were harvested and followed the same procedure of serial dilution titration for inclusion counting and further construction of one-step growth curves. These curves were also important to define the proper time-points to be used in transcriptome (RNA-seq) assays.

8.3.6. Evaluation of molecular stability of CT135 transcripts

The molecular stability of the bacterial transcripts is traditionally evaluated throughout the quantification of the mRNA half-life time after transcriptional blockage with the antibiotic rifampicin (475,476), which inhibits the *de novo* synthesis of RNA. Since the quick degradation of mRNAs from genes that are deleterious in a given environment is believed to confer selective evolutionary advantage, we sought to evaluate the molecular stability of CT135 transcripts before and after the spread of CT135-negative clones. We performed this assay for the clinical isolates with emergent CT135-negative clones. After a preliminary optimization stage, we opted to use a final rifampicin concentration of 10 µg/ml (Sigma-Aldrich) since it allowed a maximum bacterial transcriptional arrest while causing no visible morphologic effects on HeLa229 cells (data not shown). This concentration is in agreement with the one used in a previous study aiming to assess the molecular stability of transcripts in other *Chlamydia* species (476). For the assays to determine CT135 mRNA half-life time, fresh EB-enriched inocula from each isolate under passaging (for both ancestral and evolved populations) were inoculated as above on 6-well plates, and cultures were allowed to grow until antibiotic treatment. Considering that higher levels of CT135 transcripts are detected at early stages of the *C. trachomatis* life-cycle (101,324), we opted to evaluate the molecular stability of the CT135 transcripts at this stage. Hence, at 4h pi, the culture medium was replaced by fresh medium containing rifampicin and after 10 min cells were scraped with PBS and mechanically disrupted by sonication. The final suspension was rigorously divided into two identical aliquots and immediately frozen in liquid N_2. Culture samples were similarly harvested before rifampicin addition (at 4h pi). The twin aliquots were submitted to independent DNA and RNA extraction using QIAamp® DNA Mini Kit (Qiagen) and RNeasy Mini Kit (Qiagen), respectively, according to manufacturer's instructions. Purified RNA samples were further subjected to cDNA generation using TaqMan Reverse Transcription (RT) Reagents (Applied Biosystems, Life Technologies, Branchburg, NJ, USA), as previously described (251). cDNA samples were used to quantify the transcript's level through qPCR before and after rifampicin treatment. In order to overcome putative bias associated with differential degradation along the operon, we applied two independent qPCRs targeting the two genes (CT134 and CT135) (202), and calculated a mean value. The mRNA levels were normalized against the number of *C. trachomatis* genomes quantified on the corresponding DNA samples, using a previously described qPCR (231). All quantifications were performed using a LightCycler® 480 (LC480) Instrument (Roche). The reagents consisted of LightCycler® 480 SYBR Green I Master (Roche), 400 nM of each primer and 5 µl of sample DNA or cDNA, in a final volume of 25 µl. The sequences of all primers used in this study are listed in the Supplemental Table S8.2. The mRNA half-life times (t ½) were calculated by using an adaptation of the "two-fold" decay step method (477) based on the fit of an exponential decay between values obtained at the first time-point (before rifampicin addition) (N_0) and the values calculated *t* minutes (10 min) after the transcriptional arrest (N_1), using the formula: $t\,½ = -\ln 2/k$; where the rate of decay rate (k) was estimated as follows: $k = \ln(N_1/N_0)/t$.

8.3.7. Evaluation of the impact of CT135 loss on *C. trachomatis* transcriptome

Analyses of global gene expression in *C. trachomatis* through microarrays or RNA-seq have been almost exclusively performed at mid-stage of the developmental cycle in order to ensure the detection of transcripts from a higher number of genes (99-101,207,383). Here, we also focused on the bacterial mid-cycle, and chose the time-point for RNA-seq differential expression analysis according to the distinct growth dynamics observed for the D/CS637/11 CT135-positive and CT135-negative isolates (evaluated by IFU-based growth curves) (see Results section). Hence, HeLa229 confluent monolayers (cultured on T25 cm^2 flasks) were infected with serovar D CT135-positive (passage 7) and CT135-negative (passage 20) isolates at a MOI of about 1 (as described above), and were grown for 30 h and 24 h post-infection, respectively. Subsequently, the culture medium was discarded, cells were harvested using 2:1 volumes of RNAprotect bacteria reagent (Qiagen, Valencia, CA), sonicated, and subsequently centrifuged at 700 rpm for 7 min. The supernatant was collected, pelleted by centrifugation (14000 rpm for 15 min) and further processed for RNA extraction using the RNeasy Mini Kit (Qiagen), according to manufacturer's instructions. Samples were treated on column with RNase-free DNase (Qiagen), and RNA samples were subjected to Agilent Bioanalyser (Agilent Technologies) analyses before and after the bacterial mRNA enrichment using MICROBEnrich (Ambion, Austin, Texas) and MICROBExpress (Ambion) protocols, according to manufacturer's instructions. Bacterial mRNA-enriched samples were further subjected to library construction (TruSeq® Stranded mRNA sample preparation kit, Illumina) and sequencing on an Illumina MiSeq sequencer using a paired-end (2x150bp) strategy (about 15M reads were dedicated *per* sample). The quality of the raw sequence data was assessed through FastQC analysis. The sequence reads were mapped (using Bowtie2) to the D/CS637/11 genome (sequenced in this study) and the expression of each chromosome and plasmid coding sequence (CDS) was normalized as fragments *per* kilobase of CDS *per* million mapped reads (FPKM) using the Cufflinks (version 2.1.1; http://cufflinks.cbcb.umd.edu/). In order to overcome the variable efficiency of the mRNA enrichment step across samples and to improve the overall robustness of transcript abundance estimates, reads from bacterial rRNA and tRNA were selectively masked in the analyses. Differential gene expression analyses between D/CT135-positive and D/CT135-negative strains were conducted using Cuffdiff, where FPKMs and fragment counts are scaled taking the median of the geometric means of fragment counts across all libraries (geometric normalization method). This rationale is identical to the one used by DEseq) (478). The Benjamini–Hochberg procedure, which controls the false discovery rate (FDR), was applied to adjust raw *P*-values for multiple testing (479). Genes were considered to be differentially expressed when the fold change of expression exceeds the two-fold and the adjusted *P*-values were <0.05. RT-qPCR validation of the RNA-seq results was performed on a LC480 apparatus (as described above) using cDNA samples obtained from total RNA aliquots collected before depletion of rRNA and polyadenylated mRNAs. RT-qPCR data for each gene were normalized against the data obtained for the 16S rRNA transcript. Differential expression data are based on two biological replicates. In order to ensure that the observed expression differences were due specifically to the CT135 loss and not to other factors underlying laboratory propagation, we performed a similar RNA-

seq experiment on the Ia/CS190/96 isolate. This shows up as a suitable control, since this strain did not reveal any mutation in the entire genome throughout the same passaging period.

8.4. Results

8.4.1. The genome make-up of studied strains represents the major branches of the species tree

We performed experimental evolution of *C. trachomatis* ocular (C/TW-3; an historical prototype strain), epithelial-genital (D/CS637/11, E/CS1025/11, F/CS847/08 and Ia/CS190/96; clinical isolates) and LGV (L2b/CS19/08; an anorectal isolate associated with LGV-proctitis) strains. Among these, we recently described the genomes of C/TW-3 (480) and L2b/CS19/08 (473). To integrate the genetic backbone of all strains used in this work in the frame of the known species phylogeny and diversity (37), we first determined and analyzed the genome of the strains after minimal passages. We observed the phylogenetic segregation of each genome within the four major branches of the species tree, where each strain clustered with the other *C. trachomatis* strains sharing not only the tissue-specificity but also the clinical prevalence of the corresponding serovars (see Supplemental Figure S8.1A). Of note, two strains (Ia/CS190/96 and F/CS847/08) revealed clear traces of recombination involving epithelial-genital strains (see Supplemental Figure S8.1B and S8.1C), similarly to what has been extensively described in literature. Globally, the present study aiming to evaluate the dynamics underlying the *in vivo* to *in vitro* transition of *C. trachomatis* involves strains representing, not only the three disease groups (ocular, epithelial-genital and LGV), but also the four major genetic branches of the species tree.

8.4.2. Initial populations reveal some degree of allelic variation

Genomic analyses of the initial populations revealed phenomena associated with heterogeneity within DNA homopolymeric tracts located in the gene encoding the cytotoxin (CT166) and upstream from the gene CT533/*lpxC*, which encodes an essential enzyme [UDP-3-O-(R-3-hydroxymyristoyl)-GlcNAc deacetylase] in the biosynthesis of lipid A (481,482). Cytotoxin will be focused on the next section as the detected variability may underlie its functionality. Regarding the CT533/*lpxC*, for strain E/CS1025/11 we found a homopolymeric tract 36 bp upstream from the start codon (between the predicted -35 and -10 TATA boxes) with a string of 12 ‘A’ in 87% of the reads and 11 ‘A’ in 9% of the reads (4% for other ‘A’ counts) (Figure 8.1). Although a similar scenario was also observed for initial populations of the other strains, we drew attention to E/CS1025/11, since we detected a progressive increment of the 11 poly(A) tract throughout bacterial propagation (Figure 8.1 and Table 8.1).

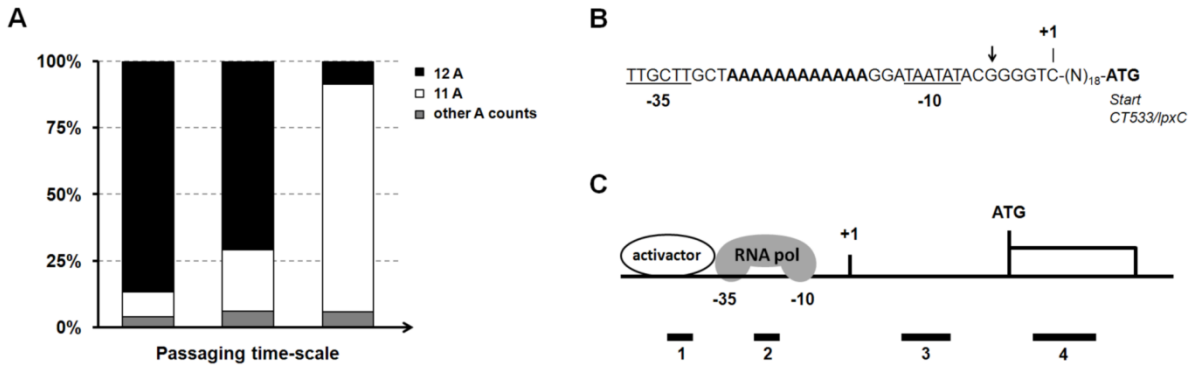


Figure 8.1. Phase variation mediated by variable homopolymeric tracts. Panel A. The graph shows the evolution throughout passaging of the percentage of sequence reads with different 'A' counts in the homopolymeric tract upstream from CT533/*lpxC* for the strain E/CS1025/11. The poly(A) tract corresponds to poly(T) in the annotated leading strand. Panel B. Schematic view of the putative promoter region of CT533/*lpxC*. The predicted transcription start site (123) is labeled by +1. The variable poly(A) tract (in bold) falls between the predicted -35 and -10 hexamers (underlined). BLAST analyses revealed the existence of variable number of 'A' counts in C. *trachomatis* genomes, and also that the nucleotide indicated with an arrow is deleted exclusively in all LGV strains. Panel C. Schematic view of the four positions (numbers 1 to 4) relative to a gene at which contingency loci (e.g., homopolymeric tracts) can cause phase variation [adapted from van der Woude MW and Bäumler AJ, Clin Microbiol Rev **17:**581-611, 2004 (395)]. Whereas positions 1 and 2 are associated with transcription initiation and position 4 with translation (ON/OFF), the mechanism regarding the position 3 is not completely disclosed. We found heterogeneity in length within homopolymeric tracts located in positions 2 (for CT533/*lpxC*), 3 (for CT043/*slc1* and the operon CT134-CT135), and 4 (for CT166 - cytotoxin).

Besides homopolymeric tracts, strains Ia/CS190/96 and L2b/CS19/08 revealed no mixture of clones, whereas the D/CS637/11, F/CS847/08 and C/TW-3 revealed allelic mixtures (Table 8.1). For the serovar D strain, the clones in the population are distinguished by two non-synonymous nucleotide polymorphisms ('GC' ↔ 'AA') affecting the same codon in the gene CT633/*hemB*, which encodes an enzyme (delta-aminolevulinic acid dehydratase) involved in the heme biosynthesis (115). A recent study showed that the heme metabolism is important for *C. trachomatis* infectivity (483), and, in *Staphylococcus aureus*, it is known that *hemB* mutants display increased ability to persist intracellularly (484). For the serovar F clinical isolate, the mixture of clones was found to involve a silent mutation in the CT682/*pbpB* gene. This gene codes for a penicillin binding protein that play a role in the final stages of the synthesis of peptidoglycan, which was recently discovered in *Chlamydia* (87). Corroborating the assumption that this mixture of clones comes from the *in vivo* population, there are *C. trachomatis pbpB* sequences in GenBank supporting both nucleotide variants. Finally, concerning the ocular prototype strain C/TW-3, the initial population contained a mixture of CT135-positive and CT135-negative clones, where the minor frequent clones carry one inactivating indel in the gene CT135 (Table 8.1). Globally, allelic mixtures were sustained by depth of coverage ranging from ~180x to ~1100x (see Supplemetal Table S8.1).

Table 8.1. Genomic alterations throughout *in vitro* passaging.

Strain	Mutation location (chromosome) [a]	Mutation location (gene) [a]	*loci* [b]	nt change [c]	aa change [d]	Frequency (passage n.º)	Gene product / Putative role
C/TW-3							
	49197-49204	---	IGR CT043/*slc1*-CT044/*ssb*	DEL: 1 bp (AA → A) [e]	---	100% (P100)	Poly(A) tract in IGR upstream of CT043/*slc1* [chaperone of T3S effectors (312,485)].
	153558	729	CT135	DEL: 1 bp (AA ↔ A)	Ψ	90%↔10% (P5-7); 100%↔0% (P100)	Putative Inc (129); Virulence gene (202,203).
	191181	28	CTW3_00885	G → A	Val → Ile	100% (P100)	Cytotoxin remnant (480)
	213502	1634	CT189/*gyrA_1*	T → C	Phe → Ser	100% (P100)	DNA gyrase subunit A (115)
	764461	1345	CT664/*cdsD* or *yscD*	G → A	Gly → Ser	100% (P100)	FHA domains-containing protein predicted to form the inner membrane ring of T3S apparatus (486).
D/CS637/11							
	152763-154214	---	CT135 [f]	DEL: 1452 bp [f]	Ψ	100% (P20-P30)	Putative Inc (129); Virulence gene (202,203).
	721797	760	CT633/*hemB*	G ↔ A	Ala ↔ Lys	80% ↔20% (P6); 0%↔100% (P20-P30)	Delta-aminolevulinic acid dehydratase (115).
	721798	761	CT633/*hemB*	C ↔ A	Ala ↔ Lys	80%↔20% (P6); 0%↔100% (P20-P30)	Delta-aminolevulinic acid dehydratase (115).
	825657	1021	CT713/*porB*	T → C	Ψ [g]	~30% (P30)	Porin (116).
E/CS1025/11							
	153379	443	CT135	DEL: 4 bp (AAATT → A)	Ψ	~40% (P20); ~100% (P30)	Putative Inc (129); Virulence gene (202,203).
	602046-602057	---	IGR CT533/*lpxC* – CT534/*lnt*	DEL: 1bp (AA↔A) [e]	---	91%↔9% (P7); 77%↔23% (P20); 15%↔85% (P30)	Poly(A) tract upstream of CT533/*lpxC* [essential enzyme in the biosynthesis of lipid A] (481,482).
	742284	2	CT645	T → C	Ψ [h]	~80% (P20); ~100% (P30)	Predicted integral membrane protein with unknown function (YGGT family) (115).
F/CS847/08							
	153373	543	CT135	C → A	Ψ	~50% (P20); ~59% (P30)	Putative Inc (129); Virulence gene (202,203).
	153742	912	CT135	INS: 2 bp (A → AAA)	Ψ	~30% (P20); ~40% (P30)	Putative Inc (129); Virulence gene (202,203).
	784264	2490	CT682/*pbpB*	G ↔ T	Leu ↔ Leu	37%↔63% (P7); 20%↔80% (P20); 44%↔56% (P30)	Penicillin-binding protein(115); Synthesis of peptidoglycan.
	826262-826265	799-802	CT713/*porB*	INS: 1 bp	Ψ	~40% (P30)	Porin (116).

				(T → TT)			
Ia/CS190/96							
	152948	315	CT135	DEL: 1 bp (AT → A)	Ψ	~28% (P30); 100% (P50-P100)	Putative Inc (129); Virulence gene (202,203).
	230974	399	CT205/*pfkA*	T → G	Phe → Leu	~96% (P100)	Diphosphate--fructose-6-phosphate 1-phosphotransferase (115).
	289987	981	CT257	DEL: 1 bp (AA → A)	Ψ	100% (P50)	CBS domain-containing protein (435) suggested to display tropism for eukaryotic lipid-droplets (318).
	825140	978	CT713/*porB*	G → A	Ψ	~100% (P100)	Porin (116).
L2b/CS19/08							
	---	---	---	---	---	---	---

[a] Locations refer to the sequences (chromosome and gene) of the high frequent clone in the first population (passages 5-7), whose chromosome and plasmid sequences were deposited in GenBank (accession numbers are listed in Table S2 in the supplemental material).

[b] The *loci* designations are based on genome annotation of the D/UW3 strain (GenBank accession number NC_000117) (115), except for the putative ORF CTW3_00885 (GenBank accession number CP006945) (480), which encodes a fragment of the ancestral *Chlamydia* cytotoxin predicted not to be functional (179). IGR, intergenic region with adjacent ORFs indicated.

[c] The nucleotide changes in open reading frames are presented in the 5' to 3' direction. DEL, deletion. INS, insertion.

[d] Mutations leading to putative protein truncation are represented by the symbol Ψ.

[e] Homopolymeric poly(A) tract upstream of the locus coding sequence corresponding to poly(T) in the annotated leading strand.

[f] The mutation event involved the entire CT135 deletion.

[g] Lost of stop codon (TAA > CAA).

[h] Mutation in the start codon: ATG (Met) > ACG (Thr).

8.4.3. Phase variation may underlie functionality of *C. trachomatis* cytotoxin

The *C. trachomatis* cytotoxin is a putative effector protein believed to act on the rapid disassembly of cytoskeleton actin filaments during the bacterial internalization process and to cause a cytopathic effect in host cells (182-184,187). Contrarily to epithelial-genital strains, LGV and oculotropic strains do not encode a functional cytotoxin (179,187). Intriguingly, we found clones in the initial populations of all epithelial-genital isolates displaying a potentially disrupted cytotoxin due to an inactivating variable 'G' count in a poly(G) tract at the 5'-end of CT166 (gene positions 29 to 37). In fact, although the predicted functional cytotoxin harbors a string of a 9 Gs, the majority of the sequence reads generated for three epithelial-genital isolates (>83% for D/CS637/11 and F/CS847/08, and >66% for Ia/CS190/96) carry a 10 'G' homopolymeric tract. This is supported by a high depth of coverage from 130x to 470x, and it is known that homopolymer-associated errors are highly reduced when using the Illumina technology (487,488), which has been used, for instance, for resequencing the variable tracts of contingency loci in *Campylobacter jejuni* (489). We believe that this variability may underlie the regulation *in vivo* of the cytotoxin functionality through an ON/OFF mechanism of phase variation. In fact, we performed a BLAST search and found out that 'G' counts other than nine are present in several recently released culture-independent *C. trachomatis* genomes (195).

8.4.4. *In vitro* passage of *C. trachomatis* results in a tropism-specific loss of the virulence gene CT135

We evaluated the emergence and spread of adaptive mutations throughout *C. trachomatis* experimental evolution (Figure 8.2). No emergent mutations were detected for the LGV-proctitis strain (L2b/CS19/08) at any time-point. In contrast, we observed inactivating mutations in the virulence gene CT135 for all the epithelial-genital strains, which rapidly rose in frequency.

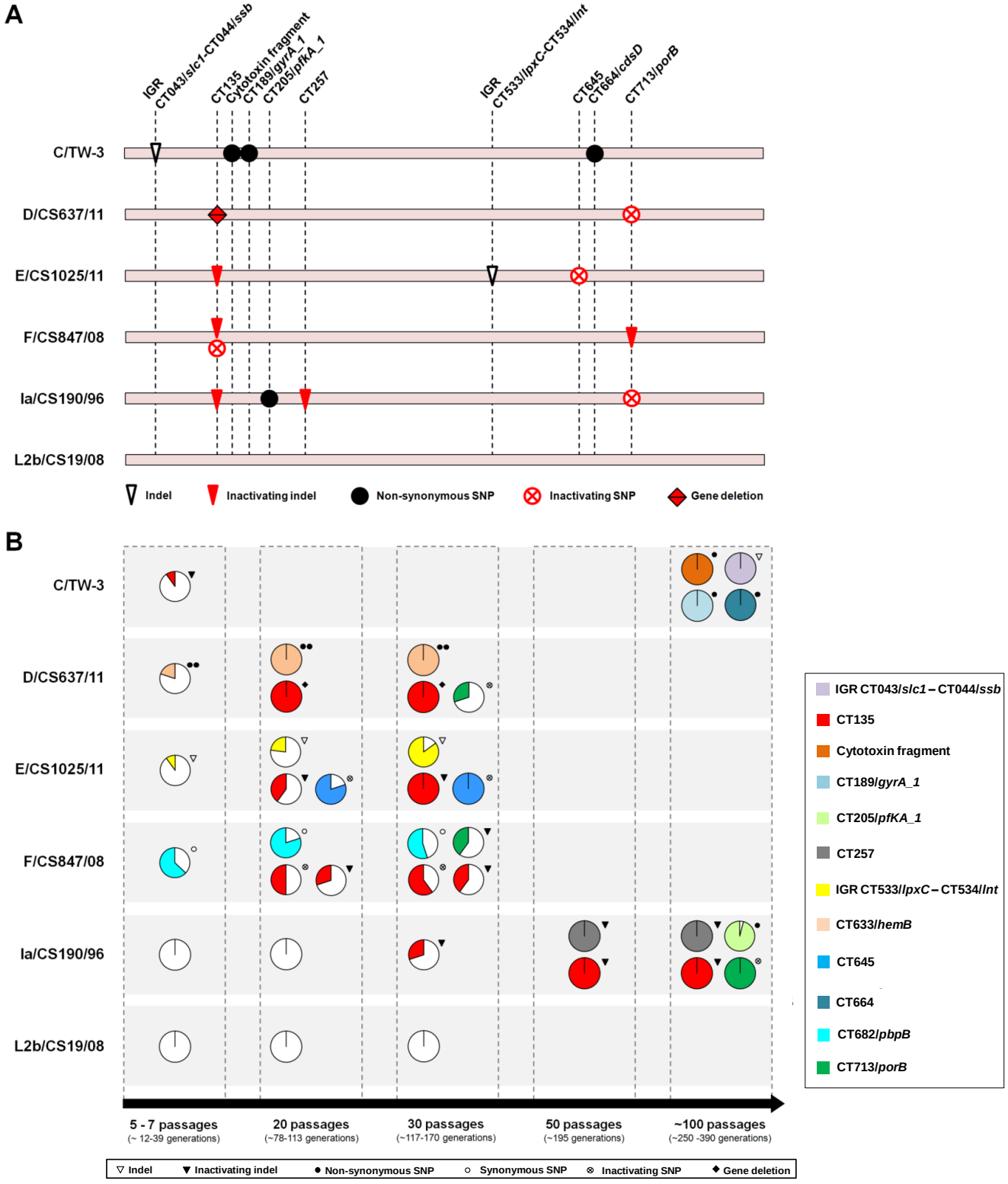


Figure 8.2. Mutational scenario throughout experimental evolution. Panel A. Chromosomal location of the genomic alterations observed during the *in vitro* passaging. The chromosomal position of each mutation (scale adjusted and given by the locus name) and the type of mutation event (inactivating events represented in red) are shown for each strain (see also Table 8.1 for details). Inactivating SNPs or indels refer to events leading to protein truncation (regardless the length of the resulting protein). For the strain D/CS637/11, the CT135 inactivating event involved the entire gene deletion between direct repeats (Figure 8.3). **Panel B**. Dynamics of the emergence and spread of mutations and their frequency in the evolving bacterial populations. For each time-point (passages 5-7, 10, 20, 30, 50 and 100), circular graphs show the frequency of the mutations in the bacterial population, where each color represents a different mutated locus. The number of bacterial generations was estimated taking into account the minimum and maximum values of the mean doubling time of the strains analyzed at each time-point, and assuming a conservative approach by considering 15 hours of exponential phase *per* bacterial life-cycle (i.e, *per* passage). Loci designations are based on genome annotation of the D/UW3 strain (GenBank accession number NC_000117).

We estimated that, in each 10 *in vitro* passages of epithelial-genital strains, a mean of 23,1 % (SD ±11,9) of the emergent clones will carry inactivating mutations in this virulence gene. An extreme example stands for D/CS637/11, where CT135-negative clones reached a frequency of 100% until passage 20. For this strain, the inactivating mutation consisted of a 1452-bp deletion between direct repeats leading to the putative formation of a fusion gene involving the flanking genes CT134 and CT136 (Figure 8.3). Three major alternative mechanisms may have mediated this event: *i)* intermolecular crossing over between DNA direct repeats; *ii)* intramolecular pairing of repeats by looping out followed by homologous recombination; and *iii)* DNA polymerase slippage during DNA replication (490-492). Of note, deletions between direct repeats have been proposed to have played a role in the genome reductive evolution of *Chlamydia* bacteria (179,493,494), namely in the cytotoxin loss in the ocular *C. trachomatis* serovars A, Ba and C (179). In support of the expression of the fusion protein CT134-CT136, we observed the existence of transcripts (in RNA-seq) compatible with this novel genomic structure. However, immunoblotting attempts to detect the protein in culture lysates were unsuccessful.

In a later point of the evolution experiment, a parallel scenario of gene inactivation occurred for CT713/*porB* gene for three out of four epithelial-genital strains (Figure 8.2 and Table 8.1). Since this parallelism provides a strong signal of metabolic adaptation of *C. trachomatis* to the culture conditions, we will discuss it in detail in the next section.

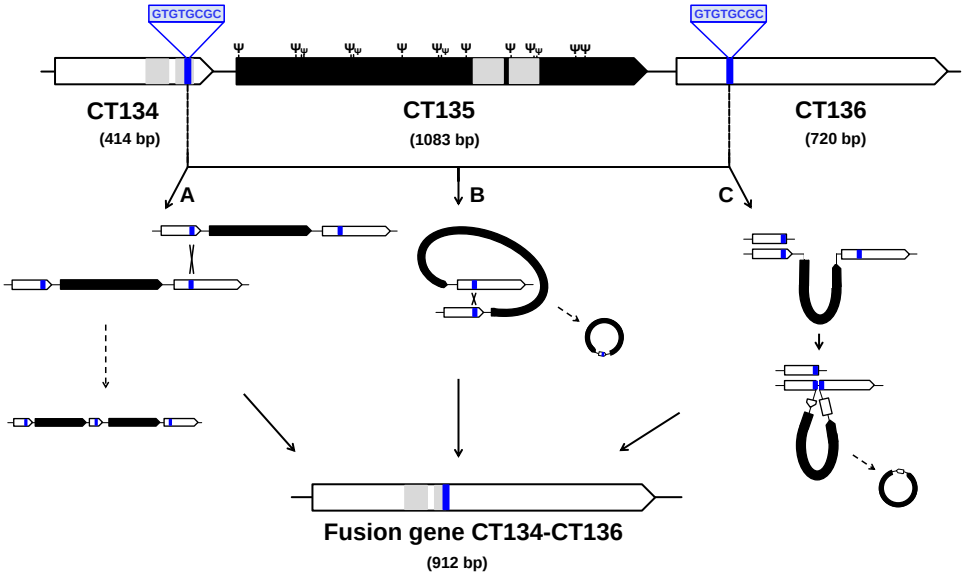


Figure 8.3. Schematic representation of the CT135 deletion in the serovar D strain. The inactivating event of CT135 involved the complete gene deletion between direct repeats (in blue) and the putative formation of a fusion gene enrolling the two flanking genes (CT134 and CT136). The underlying mechanism likely relied in one of three major pathways: A - intermolecular crossing over between direct repeats followed by recombination (yielding both a tandem duplication and a deletion); B - looping out in between direct repeats followed by recombination; and C - DNA polymerase slippage during DNA replication (490-492). The figure also shows the position of all CT135 frameshift mutations (labeled by Ψ) reported here and elsewhere (164-166,202,213,432). The bilobal hydrophobic domains that putatively enable the insertion of the CT134 and CT135 proteins into the inclusion membrane (129) are highlighted in grey.

Some other adaptive mutations were fixed in the evolving populations, although no parallelism across strains was observed for the targeted genes. For the serovar E strain, we detected the emergence and evolution to predominance of clones carrying one inactivating SNP in the start codon of CT645, which encodes a predicted integral membrane protein with unknown function (YGGT family) (115) that shows homology to a protein (YlmG) potentially involved in cell division (495,496). Also, for the Ia/CS190/96 strain, two new adaptive mutations arose: an inactivating indel for CT257 and one non-synonymous SNP for CT205/*pfkA_1*, both reaching about 100% frequency. CT257 codes for a CBS domain-containing protein (435) that was suggested to display tropism for eukaryotic lipid-droplets (318) whereas CT205/*pfkA_1* encodes a key enzyme (diphosphate-fructose-6-phosphate 1-phosphotransferase) involved in the carbohydrate metabolism (115).

Regarding the ocular strain C/TW-3, we verified that the minor frequent CT135-negative clones present in the initial population evolved to extinction, reflecting a completely opposite scenario to the one observed for the epithelial-genital strains. Additionally, we observed the further fixation of four adaptive mutations reaching 100% frequency at passage 100 (Figure 8.2 and Table 8.1). These mutations targeted the following loci: the IGR upstream from the gene CT043/*slc1* [a 1-bp deletion in a

poly(A) tract affecting the transcript sequence], the genes CT189/*gyrA_1* and CT664/*cdsD*, and a putative ORF (CTW3_00885) (480) encoding a fragment of the ancestral *Chlamydia* cytotoxin. Of note, two of these mutations might have likely interfered with T3S-mediated bacterial functions, since the gene CT043/*slc1* encodes a chaperone (Slc1) of several virulence-associated T3S effectors (312,485), whereas the gene CT664 encodes a forkhead associated (FHA) domains-containing protein (CdsD/YscD) predicted to form the inner membrane ring of T3S apparatus (486,497).

We observed no mutations in the plasmid sequence for all strains, and the number of plasmids *per* chromosome did not significantly changed throughout laboratory propagation (see Supplemental Table S8.1). Globally, the observed mutations correspond to an empirical substitution rate (we used a conservative approach by considering minimal exponential phases) ranging from 9.86×10^{-9} to 1.89×10^{-8} mutations *per* base pair *per* generation (or 0.0103 – 0.0197 mutations *per* genome *per* generation), which fits data from adaptive evolution studies in bacteria, where evolving populations are expected to fix no more than ~2.5 mutations *per* 100 generations (420,447,448).

8.4.5. Inactivating mutations in the gene CT713/*porB* may reflect metabolic adaptation to the available carbon source

Three out of the four epithelial-genital strains acquired putative inactivating mutations in CT713/*porB* that increased in frequency (Figure 8.2 and Table 8.1). CT713/*porB* gene codes for a porin that specifically enables the uptake of dicarboxylates into the tricarboxylic acid (TCA) cycle (116). This cycle is incomplete in *C. trachomatis*, and requires either exogenous glutamate (transported by the major outer membrane protein - MOMP) or 2-oxoglutarate (transported by PorB) from the host cell (115,116). Hence, as we supplemented the chamydial cultures with glutamate, the 2-oxoglutarate uptake pathway might become expendable, leading to both the inactivation of the porin-encoding gene CT713/*porB* and exclusive maintenance of the TCA feeding through the glutamate pathway (Figure 8.4). This hypothesis would rely on a typical scenario of metabolic adaptation to the available carbon source, and fits well with the postulation that the reduction of importance of the TCA cycle metabolic pathway throughout the adaptation of *Chlamydiae* to an intracellular lifestyle led to the pseudogenization of TCA cycle-related genes (498).

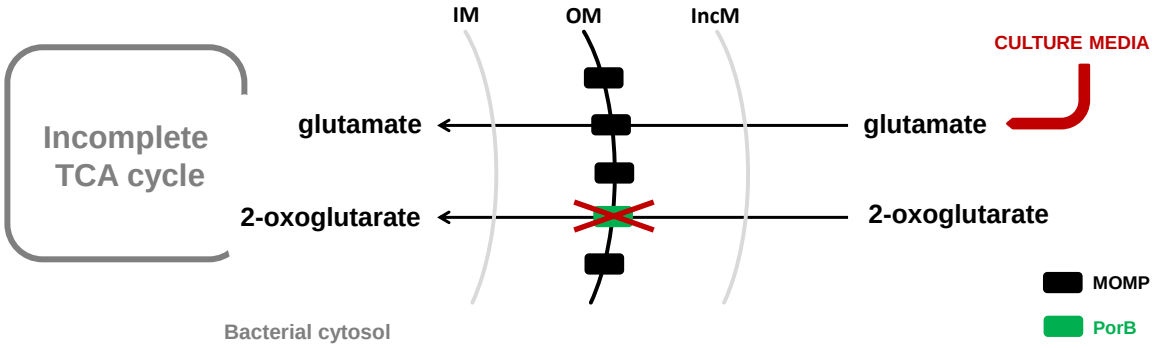
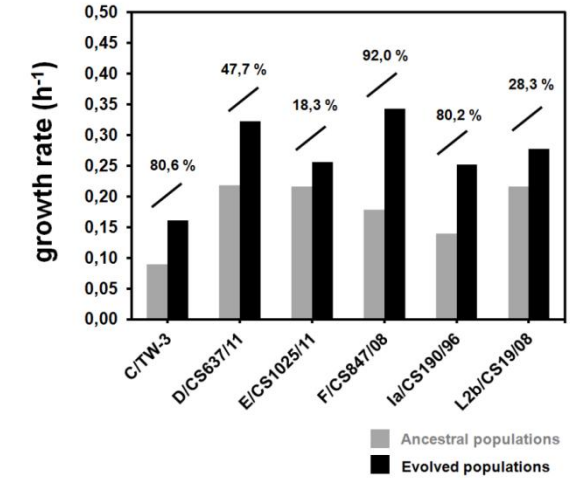


Figure 8.4. Proposed schematic representation of the adaptive process underlying the inactivation of the CT713/*porB*. The porin PorB enables the uptake of exogenous 2-oxoglutarate, whereas the general porin and more abundant Major Outer Membrane Protein (MOMP) transports glutamate. Any of these carbon sources allow *C. trachomatis* to complete the tricarboxylic acid (TCA) cycle. In the proposed model, the exclusive availability of glutamate in the culture media led the evolving bacteria to adapt towards the loss of the 2-oxoglutarate pathway through the genomic inactivation of the CT713/*porB*. IM, inner membrane. OM, outer membrane. IncM, inclusion membrane.

8.4.6. The growth rate of laboratory evolving populations increased relative to the ancestral populations leading to shorter life-cycles

In *Chlamydia*, fitness estimates essentially rely on evaluating the growth rates or IFU yield rather than on capturing the phenotype by which the adaptive advantage operates, as knock-out assays are not optimized yet for such purpose. Thus, we searched for changes in the fitness of the evolving populations by comparing the growth rates of the strains at the beginning and at the end of the study, a methodology largely applied in bacterial experimental evolution studies (451). Globally, the growth rate of all evolved populations relative to their ancestors increased from 18,3% (for E/CS1025/11) to 92,0 % (for F/CS847811) (Figure 8.5A and 8.5B), reflecting the progressively shortening of the bacterial life-cycles observed through phase-contrast microscopy. In addition, we complemented this data by constructing a one step-growth curve (i.e., enumeration of IFUs over time) to assess the length of the life-cycle before and after the CT135 loss by the serovar D strain (Figure 8.5C). This assay revealed a shorter developmental cycle for the propagated strain, which was also sustained by the observation of a similar fold change in the genome copy number between 4h pi to 30h for the D/CT135-positive strain and 4h pi to 24h for the D/CT135-negative strain (data not shown).

A



B

strain	Doubling time (h) (ancestral, evolved)	growth rate (h^{-1}) (ancestral, evolved)
C/TW-3	7,76	0,089
	4,30	**0,161**
D/CS637/11	3,18	0,218
	2,15	**0,322**
E/CS1025/11	3,20	0,217
	2,70	**0,256**
F/CS847/08	3,88	0,179
	2,02	**0,343**
Ia/CS190/96	4,96	0,140
	2,75	**0,252**
L2b/CS19/08	3,20	0,216
	2,50	**0,278**

C

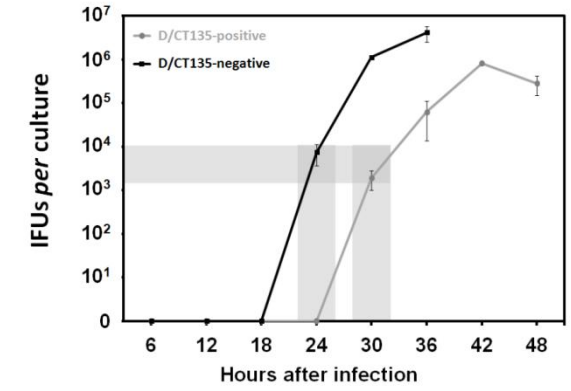


Figure 8. 5. Impact of *in vitro* passaging on the *C. trachomatis* growth kinetics. Panels A and B. Comparison of the growth rates and doubling times between ancestral (grey) and evolved populations (black). The percentage values above the bars correspond to the growth rate increment of the evolved population relatively to the ancestral. Panel C. Comparison of the one-step growth curve between D/CS637/11 CT135-positive and CT135-negative strains. Cells grown in the same conditions were infected at a MOI of 1, and cell scrapings were collected over time after infection for analysis of inclusion-forming units (IFUs). The black line represents the evolved CT135-negative D/CS637/11 strain, whereas the grey line represents the ancestor CT135-positive strain. The shaded area indicates the time points chosen for RNA-seq differential expression comparative analyses.

8.4.7. The genomic inactivation of CT135 in the epithelial-genital strains is accompanied by a quick degradation of the mRNA

The mechanisms for controlling the mRNA processing and decay play an important role in the continuous adjustment of the levels of gene expression according to the protein needs, with bacterial mRNA half-life times ranging from seconds to hours (499,500). We evaluated whether the molecular stability of CT135 transcripts was affected after protein truncation. We observed that the mRNA half-life time decreased for the epithelial-genital isolates (Figure 8.6), whereas it remained unaltered for the LGV clinical isolate. These results are concordant with an adaptive scenario where both the genomic inactivation and mRNA processing mechanisms concomitantly act to increment the competitive fitness of the CT135-negative clones. Curiously, when looking at the mRNA half-life time of the ancestral strains, we observed that the CT135-transcript is more labile for the LGV strain.

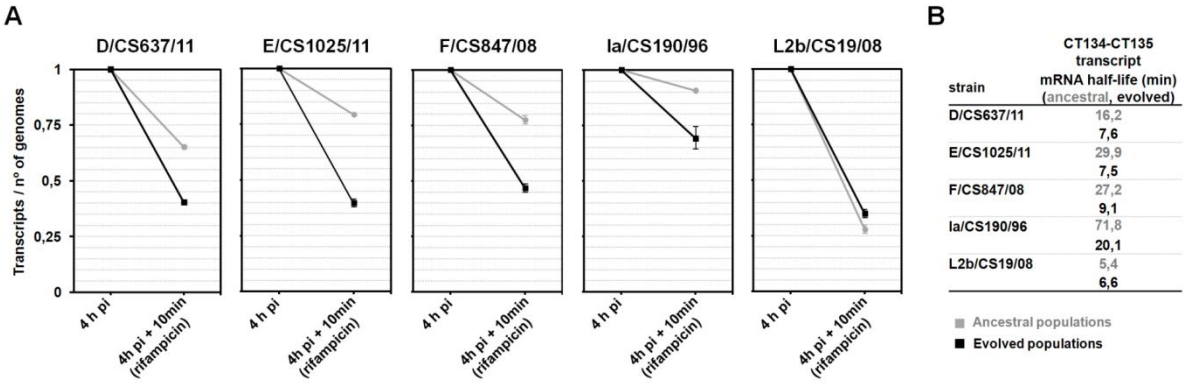


Figure 8.6. CT135 mRNA decay analysis. Panel A shows the comparison of the relative amount of transcripts at 4 h post-infection (pi) and after 10 min of transcriptional blockage with rifampicin (10 µg/ml) between the ancestral (grey) and the evolved populations (black). The assay was performed for all strains with emergent CT135-negative clones (i.e., all epithelial-genital isolates) and for the strain L2b/CS19/08 (control). The number of transcripts was quantified by independent RT-qPCR targeting the two genes of the operon CT134-CT135 (see methods for details), except for serovar D strain as the evolved population lacks CT135. Data was normalized against the number of *C. trachomatis* genomes quantified on the corresponding DNA samples. In order to facilitate the comparative analysis, the normalized value before rifampicin treatment (4h pi) was arbitrarily set to 1. Panel B shows the mRNA half-life times calculated based on the fit of an exponential decay between the quantified values at 4h pi and the values calculated 10 minutes after the transcriptional arrest.

8.4.7. Genomic deletion involving CT135 affects the expression of multiple virulence-associated genes

The mechanism by which CT135 promotes *C. trachomatis* virulence is unknown. Taking into account the rapid and parallel loss of the CT135 in all epithelial-genital strains, we wonder if the *in vitro* inactivation of CT135 is beneficial likely because it reduces the expression of several genes, and

hence the associated energetic costs of unused functions. If so, the scenario of CT135-mediated virulence would mirror the one of the plasmid-encoded Pgp4 that was shown to impact the expression of multiple genes (207). We performed RNA-seq analyses to compare the global gene expression of the serovar D CT135-positive and CT135-negative strains. Here, we are assuming that potential changes in gene expression are essentially due to the CT135 loss, although the deletion of 1452-bp also partially involved the flanking genes CT134 and CT136 (Figure 8.3). In fact, CT136 encodes a Lysophospholipase esterase, whose function as expression regulator is implausible. Regarding CT134, although it was never described to be targeted by inactivating mutations, it is predicted to belong to the same operon as CT135 (202), so a synergistic effect may be hypothesized. By using a false discovery rate cutoff of 0.05 and a fold-change cutoff of >2.0, we identified 48 significantly differentially expressed genes and one non-coding RNA, all being down-regulated in the CT135-negative strain (Table 8.2 and Figure 8.7). To confirm these results, we used RT-qPCR to evaluate the expression of a set of the highlighted genes and found a good level of correlation (slope 1.01, Pearson correlation 0.958, N = 7) (see Supplemental Table S8.2).

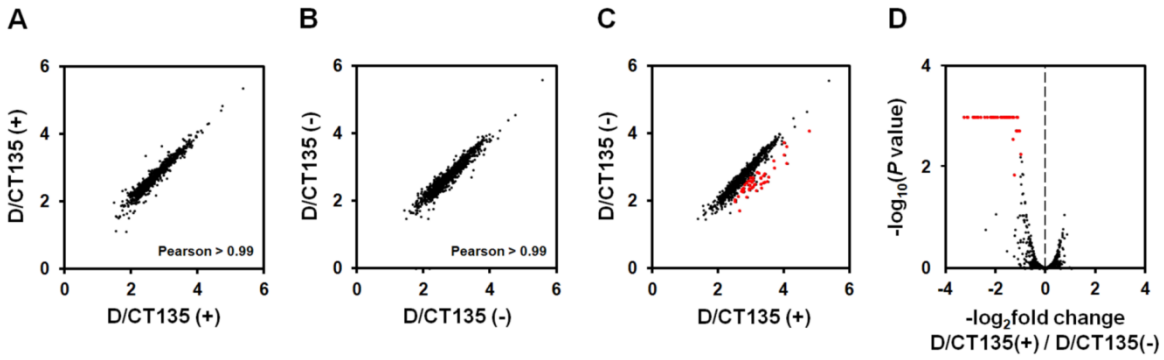


Figure 8.7. Comparative analysis of global gene expression (RNA-seq) between D/CT135-positive and D/CT135-negative populations. Panels A-B. Comparison of gene expression between biological replicates for the D/CT135-positive (A) and D/CT135-negative (B) populations. Pearson correlation coefficients are shown. Panel C. Comparison of gene expression between the D/CT135-negative and D/CT135-positive populations. The red points mark genes and the non-coding RNA for which the fold change of expression exceeds two-fold and the FDR-corrected *P*-values were below 0.05. For panels A to C, axes are $\log_{10}$-transformed normalized expression levels (FPKM). Panel D. Volcano plot of $-\log_2$ fold change (D/CT135-positive *versus* D/CT135-negative) *versus* $-\log_{10}$ adjusted *P*-values. In order to better fit the scale to data, corrected *P*-values $\leq 10^{-3}$ were set as 10^{-3}. Points in red indicate genes and the non-coding RNA for which the fold change of expression exceeds two-fold and the FDR-corrected *P*-values were below 0.05.

Remarkably, the pool of genes that are down-regulated after CT135 loss involves multiple virulence-associated genes (some of them belonging to the same operon) (Table 8.2). We highlight genes coding for: *i*) the most prominent adhesins (e.g., CT443/OmcB) (107,108) and T3S effectors (e.g., CT456/Tarp, CT694 and CT875/TepP) known to play a role in the *C. trachomatis* host-cell invasion

process (140,141,266,368); *ii*) multiple T3S-related proteins, such as substrates (e.g., CT082, CT619-20 and CT847-9) (132,290,310,312,369), a chaperone (CT576/Scc2) (501) and components of the translocation pore (e.g., CT578/CopB and CT579/CopD) (339); *iii*) genes encoding proteins putatively related to the chlamydial protease/proteasome-like activity factor (CPAF), either potential substrates (e.g., CT005, CT288, CT443/OmcB, CT456/Tarp and CT694-5) (136,502,503) or other virulence proteases acting in the same pathways in the subversion of host-cellular functions (e.g., CT441/Tsp and CT868/ChlaDub1) (134,135,269,504); and *iv*) virulence-associated proteins regulated by the plasmid-encoded Pgp4 (e.g., CT049-CT051, CT142-CT144 and CT798/GlgA) (207). Regarding the latter set of genes, we looked at the differential expression of their regulator-encoded gene (ORF6/*pgp4*) and observed a slight decrease of expression (~1.4-fold) in the CT135-negative strain. Thus, we hypothesize that the down-regulation of those genes might have been mediated by *pgp4* underexpression. Of note, the down-regulated genes include all members (CT619, CT620, CT621, CT711, and CT712) except CT621 of a family of chlamydial T3S substrates characterized by a domain of unknown function (DUF582 proteins) that are believed to target nuclear cell functions (310,317). The highly down-regulated non-coding small RNA (sRNA) (located between CT080 and CT082, and previously designated ctrR0332 in the L2b/UCH-1 strain) (Table 8.2) was previously found to be overrepresented in EBs, which is consistent with the profile found for the majority of the affected transcripts (102). Noteworthy, we observed only one differentially expressed gene (CT377/*ltuA*) in the control experiment with the Ia/CS190/96 strain. This result suggests that the short-term laboratory propagation of *C. trachomatis* does not substantially change gene expression in absence of genomic alterations.

Table 8.2. Genes and a non-coding RNA found to be down-regulated in the serovar D CT135-negative strain.

locus [a, b]	- $\log_2$[fold change D/CT135(+) / D/CT135(-)] [c]	Putative role / experimental evidence
sRNA [d]	-3,274	Abundant non-coding RNAs differentially expressed in EBs and RBs (102).
CT082 [e]	-3,153	T3S substrate (312,313). Role in host cell invasion and infectivity? (102).
CT814.1	-3,139	Putative Inc (129). Role in host cell invasion and infectivity? (102).
CT814	-3,118	Role in host cell invasion and infectivity? (102).
CT635	-2,913	---
CT579/*copD* [f]	-2,880	T3S translocation pore component (CopD) (339). Role in host cell invasion and infectivity? (102).
CT578/*copB* [f]	-2,825	T3S translocation pore component (CopB) (339). Role in host cell invasion and infectivity? (102).
CT576/*scc2* [f]	-2,762	T3S chaperone Scc2 (501). Role in host cell invasion and infectivity? (102).
CT577 [f]	-2,693	Role in host cell invasion and infectivity? (102).
CT005	-2,591	Putative Inc (156). Putative CPAF substrate (503). Role in host cell invasion and infectivity? (102).
CT080/*ltuB*	-2,438	Late transcription unit B protein (505). Role in host cell invasion and infectivity? (102).
CT444/*omcA* [f]	-2,348	Cysteine-rich outer membrane protein. Role in host cell invasion and infectivity? (102).
CT443/*omcB* [f]	-2,214	Cysteine-rich outer membrane protein (107). Adhesin (107,108). Role in host cell invasion and infectivity? (102). Putative CPAF substrate (502).
CT848 [f]	-2,180	Putative T3S substrate (132). Role in infectivity in *C. muridarum*? (431).
CT456/*tarp*	-2,162	Translocated actin-recruiting phosphoprotein (TARP) / early T3S effector prepackaged into EBs (140). Involved in host cell invasion (141). Putative CPAF substrate (503).

CT875/*tepP*	-2,152	Translocated early phosphoprotein (TepP) / early T3S effector involved in host cell invasion (368). Role in host cell invasion and infectivity? (102).
CT694 [f]	-2,140	Early T3S effector (prepackaged into EBs) involved in host cell invasion (266). Putative CPAF substrate (503).
(CT622)	-2,120	Effector (265). Putative plasmid-related virulence protein (207). Role in host cell invasion and infectivity? (102).
CT847 [f]	-2,108	T3S effector putatively involved in the modulation of the cell cycle (290).
CT181	-2,093	Role in host cell invasion and infectivity? (102).
CT046/*hctB*	-2,090	Histone-like protein 2. Mediate the DNA condensation that is characteristic of EBs (114).
CT849 [f]	-2,017	T3S substrate (369). Implicated in *C. muridarum* adherence to host cells (472).
CT620	-1,925	T3S effector (DUF582 family) believed to target nuclear functions (310,317).
CT392/*yprS*	-1,807	--
CT619	-1,738	T3S effector (DUF582 family) believed to target nuclear functions (310,317).
(CT798/*glgA*)	-1,736	Glycogen synthase. Plasmid-related virulence protein (207). Effector (321).
(CT051) [f]	-1,670	Plasmid-related virulence protein (207). Pmp-like protein (156,257).
CT441/*Tsp*	-1,651	Effector. Tail-specific protease. Subversion of host cell functions by preventing the host NF-κB activation (134,135,504).
CT868/*ChlaDub1*	-1,648	Effector with deubiquitinating and deneddylating activity (protease) (268). Subversion of host cell functions by suppressing the NF-κB activation (269).
CT356/*yyaL*	-1,631	Thioredoxin domain-containing protein (435). Role in host cell invasion and infectivity? (102).
CT695 [f]	-1,623	Early putative T3S effector prepackaged into EBs (266). Putative CPAF substrate (136).
CT288	-1,619	Putative Inc (259). T3S substrate (132). Role in host cell invasion and infectivity? (102). Putative CPAF substrate (503).
(CT565)	-1,572	Putative Inc (129). Putative plasmid-related virulence protein (379).
CT546	-1,542	Predicted outer membrane protein (506).
CT712 [f]	-1,525	Putative T3S effector (DUF582 family) believed to target nuclear functions (310,317). Putatively packaged into EBs (266).
CT365	-1,459	Putative Inc (129). T3S substrate (368). Role in host cell invasion and infectivity? (102).
(CT702)	-1,441	Putative plasmid-related virulence protein (207,379).
CT214	-1,435	Putative Inc (156). Role in host cell invasion and infectivity? (102).
CT711 [f]	-1,334	Putative T3S effector (DUF582 family) believed to target nuclear functions (310,317).
(CT049) [f]	-1,325	Pmp-like protein secreted into the inclusion lumen (156,257). Putative plasmid-related virulence protein (207).
CT837	-1,300	---
CT052/*hemN_1*	-1,275	Coproporphyrinogen III oxidase (115).
(CT144) [f, g]	-1,244	T3S substrate (369). Putative plasmid-related virulence protein (207).
(CT142) [f, g]	-1,180	T3S effector (369). Putative plasmid-related virulence protein (207).
CT372/*oprB*	-1,155	Outer membrane protein with a carbohydrate-selective porin (OprB) (507).
(CT050) [f]	-1,125	Pmp-like protein secreted into the inclusion lumen (156,257). Putative plasmid-related virulence protein (207).
CT564/*yysT*	-1,114	T3SS Integral membrane structural protein (Yop proteins translocation protein T) (291).
CT735/*dagA_2*	-1,037	Na(+)-linked D-alanine glycine permease (115).
CT659	-1,010	Predicted RNA binding protein (99). Role in host cell invasion and infectivity? (102).

[a] Loci nomenclature and numbering refer essentially to the annotated genome from the D/UW3 strain (GenBank accession number NC_000117).

[b] Loci in bold are potentially down-regulated on behalf of the CT135 loss. Non-bold loci (also in parentheses) were previously found to be transcriptionally regulated by the plasmid (namely by the plasmid-encoded protein Pgp4) (207,379). In the present study, the down-regulation of this latter set of genes may be associated with Pgp4 as it was found to display lower expression levels in the D/CS637/11 CT135-negative population.

[c] Adjusted P-values < 0.05 and a fold-change cutoff > 2. Genes are ordered according the magnitude of differential expression.

[d] Refers to the previously identified sRNA ctrR0332 (102) that actually comprehends two sRNAs (likely processed from a larger transcript). In the D/UW-3 annotation, it is located inside a non-existing but previously annotated ORF (CT081). Both sRNA were similarly down-regulated.

[e] CT082 was suggested to be up-regulated in the *pgp4* knockout mutants (207). As in the present study it was found to be strongly down-regulated even with a slight *pgp4* expression decrease, we assumed this result as a consequence of CT135 loss.

[f] These genes are likely expressed in the same transcriptional unit as other down-regulated contiguous genes (also marked).

[g] Although it is believed that the CT143 gene is coordinately expressed with the flanking genes CT142 and CT144 (369), the obtained differential expression value was slightly below the cutoff.

8.5. Discussion

It is known that pathogens adapt to laboratory culture conditions, whereby *in vitro*-maintained strains may no longer reflect the circulating isolates. On behalf of this, the use of laboratory-passaged strains has been frequently questioned in *Chlamydia* research (32,195,198-201). Although the adaptive process behind the *in vivo* to *in vitro* transition is not understood, cumulative data have pointed out that a deep characterization of this process is mandatory, since: *i*) mixed-clone populations have been found in multiple culture stocks from both *C. trachomatis* and *C. muridarum* strains, with clones displaying distinct *in vivo* virulence (32,202,431,508); *ii*) prototype strains have been suggested to display distinct phenotypic presentations when comparing with recent clinical isolates, such as distinct growth-rates and/or yielded progeny (199,434) and *in vivo* infectivity capacities (200); *iii*) *in vitro* passage selected for *C. muridarum* displaying severely attenuated *in vivo* pathogenicity (472); *iv*) clones with frameshift mutations in the *C. trachomatis* virulence gene CT135 (or in the *C. muridarum* homolog TC_0412) have been detected in several cultured populations (32,201,202,213,350,432,508); and *v*) the use of cell-culture propagation in *Chlamydia* research will expectedly be boosted due to the recent development of transformation and gene knockout experiments in *Chlamydia* (207,213-217,219,509,510).

In the present study, we specifically aimed to assess the genomic and transcriptomic dynamics of *C. trachomatis* after its introduction in culture. As the experimental design mirrored the typical culture techniques, i.e., assisted bacterial entry and exit (by centrifugation and sonication, respectively), which discards the selective pressure underlying these steps, no discussion will be performed about evolutionary adaptation targeting specifically the bacterial attachment and host-cell lysis. This would require a study with different design and goals (472). We observed a dynamics of mutation fixation for this obligate intracellular bacterium that clearly fits the typical scenario of adaptive evolution seen in similar experiments in extracellular bacteria (448,463). The most frequently observed events were inactivating mutations followed by non-synonymous mutations (none of the emergent mutations in ORFs were synonymous), and the mutant clones progressively rose in frequency in the population, clearly supporting that mutations were beneficial (Figure 8.2). In general, our results show an adaptive scenario underlying the *in vivo* to *in vitro* transition for epithelial-genital strains that seems to be essentially marked by two major stages within different time-frames: *i*) a rapid and parallel evolutionary loss of the virulence gene CT135 (and the consequent down-regulation of multiple CT135-related virulence genes); and *ii*) a later inactivation of the porin PorB (encoded by the gene CT713/*porB*), which may be a result of a metabolic adaptation to the carbon source (glutamate instead of 2-oxoglutarate). Regarding the latter, it is intriguing why not all strains were affected, which we speculate to be a matter of time-scale. Future experiments involving the switch between glutamate and 2-oxoglutarate coupled with the analyses of the population dynamics will certainly dissect this hypothesis. These results provide an explanation to the enigmatic emergence of both CT135-null (202,213,350,432) and *porB*-null mutants (511) in *in vitro* populations. Notably, CT135-negative clones rapidly outcompeted CT135-positive clones, suggesting a strong selective force against CT135 functionality. Although we cannot exclude that inter-clone recombination may have also contributed to

this notable frequency increase, values of this magnitude were only previously observed under strong selective pressures with antichlamydial compounds (512).

CT135 encodes a putative Inc (129) that contains a T3S signal recognized by heterologous bacteria (324), but whose function is unknown. We found that CT135 is a direct or indirect virulence regulator, as its loss implicated a down-regulation of multiple genes, where more than 30 have been previously associated with *C. trachomatis* virulence properties (Table 8.2). We anticipate that CT135 loss and downstream down-regulation of virulence genes constitutes the mechanism behind the observation that CT135-null mutants showed attenuated *in vivo* infections in mice (202,203). In support of this, an interesting recent study from Ramsey's group (201) showed that when mice were inoculated with *C. muridarum* population displaying a mixture of positive and negative clones of TC_0412 (CT135 homolog), the former shift to predominance. Although it is believed that CT135 is an Inc protein, its location in the inclusion membrane was never demonstrated by immunofluorescence microscopy. Yet, CT135 is most expressed at beginning of the *C. trachomatis* life-cycle (101,324), coincident with the formation of the inclusion membrane, and belongs to the same operon and displays the same expression levels and profile as CT134, which also encode a putative Inc protein (129,202,324). As the the C-terminal region of CT134 was also lost in the deletion event, we could also speculate the existence of a cooperative action involving CT134 and CT135, where CT135 effect would be boosted by the "upstream" protein.

Almost all affected virulence factors seem to be coordinately expressed at late or immediately early stages of *C. trachomatis* life-cycle (99-102,324,369), suggesting their packaging into EBs for subsequent rounds of infection. Curiously, some of them were previously shown to be transcriptionally regulated by the well-described EUO (e.g., CT443/*omcB*, CT080/*ltuB*, CT441/*Tsp* and CT576/*scc2*), which similarly to CT135 is also an early protein (119,120), or by the plasmid Pgp4 (e.g., CT049-CT051, CT142-CT144, CT622 and CT798/*glgA*) (207). Whereas CT446/*euo* revealed no changes in expression, the slight lower expression of *pgp4* in the serovar D CT135-negative strain may justify the detected down-regulation of Pgp4-targets. This latter observation leaves open the possibility that the attenuation of *C. trachomatis* virulence *in vitro* may occur by more than one mechanism (such as genomic inactivation of CT135 and/or down-regulation of the virulence regulator Pgp4), which potentially take place in a cumulative or differential fashion depending on the strain. Future transcriptional analyses of other *in vitro* passaged strains, namely CT135-unaltered strains and strains with an extent passaging time-scale, will certainly be important to confirm this hypothesis.

The most down-regulated locus (about 10-fold) was the previously described sRNA ctrR0332 (102). Although increasing data have shown that the expression of sRNAs may be determinant for virulence in both extracellular (513) and intracellular bacterial pathogens (514) and, in *C. trachomatis*, one sRNA (IhtA) was found to control the timing of RB to EB transition (117,118), the specific role of the sRNA ctrR0332 on *C. trachomatis* virulence cannot be determined yet. However, it is tempting to hypothesize that it may regulate CT135 (for instance, by controlling its mRNA stability), whose loss would render the sRNA dispensability. Collectively, all these data indicate that the control of *C. trachomatis* virulence at the stage where EBs are loaded with virulence determinants is likely multi-factorial.

The scenario of the rapid loss of CT135 in the transition from *in vivo* to *in vitro* raises relevant issues concerning studies where intensive laboratory propagation is necessary, such as mutagenesis, transformation and/or drug testing. In fact, these may enhance the selection of CT135-negative mutants, where a single inactivating event may have a downstream strong impact on the virulence and pathogenicity of the strains, introducing bias when interpreting data on the course of their use. For example, the loss of virulence after *in vitro* propagation is well-documented for other important human pathogens, such as *Mycobacterium tuberculosis* (469), *Staphylococcus aureus* (470), *Coxiella burnetii* (462) and *Campylobacter jejuni* (515). Therefore, it seems important that researchers performing studies with laboratory-propagated *C. trachomatis* epithelial-genital strains confirm the integrity of CT135 in selected clones. This issue is also raised when working with *C. muridarum*, as the interpretation of the results from recent studies searching for targets of antibiotic resistance (516) or evaluating pathogenicity *in vivo* (472) was complicated by the presence of frameshift mutation in CT135-homolog.

One intriguing result was the observation that CT135 inactivation occurred for all epithelial-genital strains (regardless of both their genetic backbone and the clinical prevalence of their representative serovars), but not for strains with different tissue tropism (LGV and ocular strains). Considering that *C. trachomatis* is likely undergoing a directional evolution towards niche-specific adaptation to each infected tissue – ocular conjunctivae, genital mucosa and lymph nodes (37,163,173,301), one might speculate that CT135 may be a direct or indirect determinant of tissue tropism, as suggested for other *C. trachomatis* loci (e.g., the cytotoxin gene and the tryptophan operon) (51,167,179). Several lines of evidence sustain that the LGV strains evolve to retain the CT135 activity *in vitro*: *i*) the passaging of the LGV-proctitis associated clinical isolate do not lead to the CT135 disruption (Figure 8.2); *ii*) CT135-null mutants did not arise even when the historical LGV prototype strain (L2/434/Bu) was continuously propagated in HeLa cells during over one year (350); and *iii*) no CT135 polymorphisms were found within the population of any prototype strain from different LGV serovars (L1, L2 and L3) (202). Regarding the ocular strain that we have used (C/TW-3), we observed a scenario where a sub-population of CT135-null clones evolved to extinction (Figure 8.2 and Table 8.1), which is enigmatic since the CT135 was already found to be disrupted in several annotated genomes of ocular strains (37,166,213,517). Considering all these observations, and although the HeLa cell line is the most commonly used for culturing *C. trachomatis*, we wonder if a similar scenario is obtained by propagating LGV and ocular strains in tropism-related cell-lines, such as monocytic or conjunctival cell lines, respectively. Although the CT135 transcripts have been found to be more labile for the LGV strain (Figure 8.6), we cannot discard its contribution to the ability of *C. trachomatis* LGV strains to infect macrophages and disseminate to lymph nodes. In fact, the CT135 protein sequence has five LGV-specific amino acids (data not shown), and was previously found to be among a pool of genes likely involved in phenotypic differences among LGV strains (473).

The growth rate of all laboratory evolving populations increased relative to the ancestral populations (leading to shorter life-cycles), reflecting a gradual Darwinian fitness improvement over time (Figure 8.5). Although we cannot attribute the life-cycle shortening to the inactivation of CT135 (since the ocular and the L2b strains also improved their fitness, and other mutations arose), it is

noteworthy that *in vivo* studies also suggested that CT135 operates in sustaining a more persistent or chronic infections in the female mouse genital model (203). Other fixed mutations (Figure 8.2 and Table 8.1) have also the potential for altering the bacterial fitness likely by affecting: *i*) the cell division (mutation in CT645); *ii*) the nutrient acquisition/processing (mutations in CT205/*pfkA_1*, CT257 and CT713/*porB*); or *iii*) global regulatory mechanisms, such as DNA supercoiling (mutation in CT189/*gyrA_1*). Finally, despite most gene targets showed no parallelism across strains (at least for the time-scale evaluated), we cannot discard that these mutational events may be related to the CT135 loss [(either functionally or genetically (e.g., hitchhiking)].

In the present experimental evolution study, we also found genetic features potentially underlying phase variation mechanisms in *C. trachomatis*. Phase variation may be driven by reversible alterations in the genotype that causes either frameshift or mutations in non-coding regions impacting gene expression (Figure 8.1), and has been associated with genes involved in host-pathogen interactions (393,395). Within *Chlamydiaceae*, to our knowledge, phase variation mediated by homopolymeric tracts was only hypothesized to occur in *C. pneumoniae*, affecting members of the paralogous families Pmps (396,397) and Cpn 1054 (518). Here, we detected variable homopolymeric tracts that have the potential to modulate the functionality of genes encoding proteins likely involved in *C. trachomatis* pathogenesis: a poly(G) tract within CT166 [encodes the cytotoxin (187)] and a poly(A) tract in the promoter region of CT533/*lpxC* [encoding a critical enzyme in the biosynthesis of lipid A of lipopolysaccharide (LPS) (482)]. Regarding the cytotoxin, variable 'G' counts yielding an ON/OFF protein (also present in genomes obtained directly from clinical swabs) suggest the putative existence of cytotoxin-mediated phenotypic diversity within *in vivo C. trachomatis* populations of epithelial-genital strains. Still, we wondered if the frameshift mutation in the poly(G) tract could lead to the production of an alternative shorter protein (still including the UDP-glucose binding and glycosiltransferase domains), since the translation could eventually be reinitiated at an alternative putative start codon at position 93 (as frequently annotated in *C. trachomatis* genomes deposited in GenBank). However, all RNA-seq analyses revealed no transcripts, pointing out that this short protein version is not expressed (data not shown), which reinforces our assumption that the cytotoxin regulation is carried out through a classical ON/OFF mechanism. Besides its known activity as an effector in disassembling the host cytoskeleton upon bacterial invasion (182,183,187) and it is putative role in subversion of other cellular functions (184), the chlamydial cytotoxin was proposed to account for the resistance to the anti-bacterial effect of IFN-γ (186,495,519,520). In particular, the *C. trachomatis* cytotoxin was hypothesized to target human guanylate binding proteins (hGBPs) that are potentiators of the IFN-γ activity (i.e., depletion of the essential amino acid tryptophan) (521,522). Therefore, although epithelial-genital strains are capable of using indole for tryptophan biosynthesis (146), we hypothesize that this putative phase variation mechanism may contribute for the quick adaptation of *C. trachomatis* to changes in the tryptophan availability by targeting IFN-γ-inducible hGBPs. Concerning the poly(A) tract upstream from CT533/*lpxC*, we observed an increase in frequency of clones carrying dissimilar base counts (Figure 8.1 and Table 8.1), supporting selective advantage. It affects the highly conserved length of the spacer region between the predicted -35 and -10 'TATA' boxes (Figure 8.1), which might alter the binding of the RNA polymerase to the promoter, and thus, the gene expression.

We also found this polymorphism in GenBank annotated genomes. Curiously, in a pioneer study back to 1990s (523), it was proposed that the *Chlamydia* LPS is phase-variable, since LPS phenotypic variation was found within a population in tissue culture. In other bacteria, including intracellular organisms, phase variation mechanisms affecting the LPS biosynthesis are well-documented to be associated with antigenic variation (524). For instance, it was demonstrated that a phase variation-mediated antigenic shift of LPS modulates the ability of *Francisella tularensis* to grow within macrophages (525). In this regard, a phase variation mechanism targeting the CT533/*lpxC* in *C. trachomatis* may control the biosynthesis of lipid A, and ultimately LPS antigenic variation mechanisms for evading the host immune system. Curiously, we noted that the CT134-CT135 transcript has also a poly(T) located between the TSS and the start codon of CT134 (Figure 8.1). Although no assumption can be done on its phenotypic impact, a GenBank search revealed a rather intriguing profile of nine, eight and seven 'T' counts for most LGV strains, ocular and epithelial-genital strains, respectively. Globally, although the existence of heterogeneity within homopolymeric tracts in *C. trachomatis* populations seem to be more common than previously expected and to potentially modulate important bacterial functions, their role in mediating phase variable phenotypes with impact on pathogenesis still need experimental confirmation.

As concluding remarks, our study focused on the *C. trachomatis in vivo* to *in vitro* transition reveals not yet identified putative phase variation mechanisms targeting genes involved in pathogenesis, and also potentially discloses the molecular basis underlying the CT135-mediated virulence. CT135 may be a master player in pre-loading EBs with virulence effectors required for the invasion and subversion of epithelial cells. More, as the length of time in the *in vitro* culture is a predisposing factor that contributes to the loss of this virulence gene in epithelial-genital strains, this may constitute an issue as studies demanding culture propagation might enhance the selection of CT135-negative mutants, originating less virulent strains. Thus, restricting laboratory propagation to minimal passages and evaluating the status of the CT135 genotype might constitute good premises towards both a better interpretation of phenotypic data and inter-laboratory comparisons.

Acknowledgements

VB was a recipient of Ph.D. fellowship (SFRH/BD/68527/2010) from Fundação para a Ciência e a Tecnologia (FCT). MP was funded by a Wellcome Trust ISSF grant (097831/Z/11/Z). RF and AN were recipients of a Ph.D. (SFRH/BD/68532/2010) and post-doctoral fellowships (SFRH/BPD/75295/2010) from FCT, respectively.

Chapter IX

Final overview, concluding remarks and future directions

9. Final overview, concluding remarks and future directions

In his extraordinary book "*On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*" (1859) (526), Charles Darwin launched the scientific concept that populations progressively evolve through a process of natural selection, in which individuals more fit in a particular environment become more common in the population. Darwin knew that his theory of evolution by natural selection would imply the need of "natural variation" (as he called it), in which selection would work on. This assumption would forever change the vision about the forces driving evolution, and would constitute the foundation for the current discipline of evolutionary biology. In fact, after more than 150 years, in the genomics era, we now know that random events called mutations are the ultimate source of variation, while selection for mutations with beneficial effects provides the direction in the evolution, in a concept called adaptive evolution. In this context, studying the properties and dynamics of fixation of mutations in populations of bacterial pathogens has been crucial for the understanding of the molecular basis behind adaptation, and consequently, behind specific phenotypic traits, such as resistance to antibiotics, virulence or specific preference to particular hosts or tissues. The genomic changes of microbial pathogens are intrinsically linked to the evolutionary host-pathogen arms race that takes place during the infectious process. It is notable that, after so much time, the hypotheses grounded by Darwin still inspire hundreds of research questions, including the ones that governed this Ph.D. study. In fact, the major human pathogen *C. trachomatis* exhibit a remarkable capacity of adaptation, as revealed by its highly specialized intracellular life-style and its ability to infect such distinct human tissues, constituting a good study model for testing the Darwin's concepts. By taking advantage of genomic and transcriptomic analyses, the major goals of the present thesis were: *i*) to contribute for the clarification of the evolutionary patterns underlying the adaptation of *C. trachomatis* to the different human tissues (eyes, genitalia and lymph nodes) (topic essentially addressed in chapters II to VI); and *ii*) to get insight into the adaptive process behind the *C. trachomatis in vivo* to *in vitro* transition (chapters VII and VIII). For this purpose, we developed a set of studies oriented towards the evolutionary dynamics of *C. trachomatis* genome and the molecular patterns of genes encoding proteins with unknown function.

In chapter II, we performed comparative genomics to study the evolutionary patterns underlying *C. trachomatis* speciation (divergence) and serovar radiation (polymorphism), and used robust bioinformatics tests over a set of intra-species polymorphic genes as a means to investigate positive selection events driving phenotype dissimilarities. The results show that the nonsynonymous changes likely increased their weight relative to synonymous changes since *C. trachomatis* was established as species, reflecting an evolutionary pathway of *C. trachomatis* towards the adaptation to the human being. This was reinforced by the observation of a non-random distribution of both nonsynonymous and synonymous mutations according to the dissimilar tropism and ecological success of the strains. Hence, the mutational dynamics of *C. trachomatis* strongly points that this pathogen is undergoing a directional evolution towards niche-specific adaptation, as further demonstrated by the detection of several genes (and codons) under positive selection. These include genes that may be implicated in the ability of *C. trachomatis* to infect a particular human cell-type, and

also genes that likely mediate pathogenic (and/or ecological success) dissimilarities among strains displaying the same tissue tropism. These results may contribute for prioritizing the research questions of functional studies aiming to clarify the determinants of tissue tropism, virulence or pathogenic dissimilarities among *C. trachomatis* strains. Another conclusion from the chapter II came from the observation that genes encoding proteins that directly exposed to host (mostly Inc proteins or effectors) are overrepresented by nonsynonymous mutations, being thus the ones likely more prone to be involved in pathoadaptation. This points out that the major targets of selection during the *C. trachomatis* adaptive evolution are likely proteins mediating the complex cascade of host-pathogen interactions. On behalf of this, we further carried out two studies focused on potential host-interacting proteins by: *i*) investigating whether the mutational and transcriptional profiles of *inc* genes could be related to the type of infections or tissue tropism associated with *C. trachomatis* (Chapter III) and *ii*) searching for novel *C. trachomatis* T3S substrates/effectors (performed by the L.J. Mota's group), and determining the correlation of their expression with the different stages of the developmental cycle and the distinct tropism of *C. trachomatis* strains (Chapter IV).

Among the findings from Chapter III, we highlight that specific amino acids in a subset of Inc proteins, or the differential expression of some *inc* genes, could be determinants of the unique capacity of *C. trachomatis* LGV strains to infect macrophages and disseminate into lymph nodes. We also found 19 novel *C. trachomatis* T3S substrates (including 14 putative Inc proteins), and determined the expression levels and profiles of the 48 selected *inc* genes (most of them with unknown function) throughout the developmental cycle of *C. trachomatis* of strains representing the three *C. trachomatis* disease-groups. The results from next chapter (Chapter IV) provided further support that genes encoding proteins with unknown function play key roles in the pathobiology of *C. trachomatis*. In fact, we found that 10 poorly characterized proteins are likely T3S substrates of *C. trachomatis*, where seven of them were delivered by *Y. enterocolitica* into host cells, further suggesting that they could be novel effectors of *C. trachomatis*. Considering the recognized role of the T3S system in the *C. trachomatis* biology, the identification and characterization of the molecular patthers of *C. trachomatis* T3S effectors has been a crucial step towards a comprehensive understanding of how *C. trachomatis* interacts and exploits the host cell for promoting its growth and, ultimately, virulence. The subsequent characterization of the expression profiles of their coding genes throughout the *C. trachomatis* developmental cycle revealed that the novel putative effectors possibly act at different stages (early, mid-cycle, late, or even constitutively throughout the cycle), which immediately raises the hypothesis that they may be involved in distinct bacteria-host interactions. In parallel to what we detected for *inc* genes (Chapter III), we additionally observed that some genes encoding the novel T3S effectors present dissimilar expression levels between *C. trachomatis* strains displaying distinct tissue tropism, and found genetic dissimilarities in promoter regions that may justify the putative disease group-specific differential expression. Considering the progress on *Chlamydia* transformation and genetic manipulation, our work highlights proteins that may be prioritized in studies applying those tools, in order to disclose either their biological role or their potential contribution to the pathogenesis and distinct tissue tropism of *C. trachomatis.*

Besides the continuous efforts to understand the virulence factors that decide the differential host-cell appetence of *C. trachomatis* strains, the current research studies in this pathogen are increasingly drawing attention to the heterogeneity among strains from the same disease-group. In fact, the genetic basis underlying pathogenic and ecological success dissimilarities among same-niche infecting strains is far from being deciphered. Hence, attempts to decode those phenotypic differences should also focus on deeply characterizing the genetic variability among those strains, where serovar-specific or even strain-specific mutations cannot be neglected. In this regard, we developed two studies (Chapters V and VI) whose major goal was to contribute for the knowledge on heterogeneity among strains from the same disease-group. In Chapter V, this was achieved by means of the fully-sequencing, annotation and publicly release of a genome of an additional trachoma-causing serovar (strain C/TW-3). A global examination of the genetic backbone of the C/TW-3 strain allowed: *i*) the observation that genetic diversity among ocular strains is likely higher than predicted, where the full-genome availability of several other ocular strains will be mandatory for the understanding the complete picture; and *ii*) the identification of some noteworthy genetic features, including potential "trachoma-specific" pseudogenes (i.e., genes potentially strictly required for functions other than those needed when *C. trachomatis* infects the ocular conjunctiva). In Chapter VI, we focused on deeply characterize the genetic variability among LGV strains and to get insight into the mutational events that occurred throughout the pathoadaptation of the epidemic L2b lineage. The major force driving this investigation was the expectation that the differential disease severity and tissue tropism among LGV strains should in part rely on the genetic backbone of the infecting strain, where genes overrepresented by intra-LGV amino acid variant sites should be determinant factors for the phenotypic dissimilarities. We were also particularly interested in searching for L2b-specific genetic features, as these could shed light on the explanation why the epidemic L2b strains apparently exhibit wider tropism and transmission skills than strains from the remainder LGV serovars. Corroborating the results from the chapters II and III, we found that effectors and/or T3S substrates (namely several Inc proteins) and other virulence determinants (e.g. plasmid-regulated virulence factors) have also been the major targets of putative adaptive changes in the evolutionary pathway leading to the divergence within the LGV biovar. Of note, as observed for the LGV-specific amino acid changes in Inc proteins (Chapter II), the large majority of intra-LGV non-silent mutations in *inc* genes affected protein regions predicted to face the cytosol of the host, once again indicating that the molecular signatures evocating adaptation are strongly linked to complex host-pathogens interactions. From chapter VI, we further observed that, among the multiple serovar-specific variant sites detected, only the L2b-specific mutations are overrepresented by amino acid changes, fitting well with the supposition that the L2b strains are the most adapted LGV strains, as revealed by both their recurrent ability to infect two distinct human cells (epithelial cells from the rectal mucosa and macrophages) and their higher clinical prevalence according to the worldwide LGV epidemiological picture. Under such hypothesis, we further get insight in the evolutionary trends throughout the expansion of the epidemic L2b isolates, and noticed that nine in each 10 of the few intra-L2b variant sites are nonsynonymous. In summary, these data contribute to elucidate both the pathogenesis of LGV and also the pathogenic and

ecological success dissimilarities among the only C. *trachomatis* strains with ability to infect the lymph nodes.

Throughout this conclusive section, it is not our intention to provide an exhaustive description and enumeration of the studied genes. However, it is worth noting that the identification of molecular profiles of hypothetical protein coding genes was cumulative and crossed the whole thesis. An example of the contribution of the present work to the clarification of the role of a gene encoding a protein with unknown function stands for CT105. In fact, CT105: *i*) encodes a T3S substrate and putative effector (Chapter IV); *ii*) is a pseudogene for ocular strains (including the C/TW3 strain) (Chapters II and V), while it shows an overrepresentation of amino acid alterations between strains from LGV and genital disease-groups (Chapter II); *iii*) shows an early-cycle profile of gene expression and higher expression in the LGV strain than in the ocular and genital strains (Chapter IV); *iv*) has a LGV-specific 74-bp insertion within the putative promoter region that may justify its potential disease group-specific differential expression (Chapter IV); and, finally, *v*) is among the *C. trachomatis* genes displaying a high overrepresentation of nonsynonymous variant sites among LGV strains (Chapter VI). Besides CT105, we have highlighted several others throughout the thesis that may constitute promising targets for functional studies aiming to decipher the role of these poorly characterized proteins in the *C. trachomatis* pathobiology.

In Microbiology, there is the general assumption that the laboratory maintenance of bacteria impacts their evolutionary pathway or phenotypic signatures through the induction of genomic and/or transcriptomic changes. Nevertheless, few data are available for obligate intracellular bacteria, largely as a result of the difficulties associated with both the need of using host cell lines to culture those organisms and the establishment of straightforward genotype-phenotype associations. In the *Chlamydia* research field, this issue has been particularly controversial because most research works have been carried out by using “prototype” strains without the knowledge of the adaptive process underlying the *in vivo* to *in vitro* transition and the long-term propagation of *C. trachomatis*. The second main objective of this Ph.D. study was to shed light in this process. In Chapter VII, we performed an unprecedented long-term (about one year) continuous serial *in vitro* propagation of *C. trachomatis*, which likely represents decades of cell culture work enrolling a specific strain of a common chlamydial research laboratory. After comparing the full-genome sequences of strains before and after their long-term propagation, we were able to conclude that *in vivo*-derived genetic make-up of a *C. trachomatis* strain is not significantly compromised by its long-term laboratory propagation, as surprisingly few mutations seem to be fixed in experimental evolving populations. This directly implies that results from comparative genomic studies enrolling the "multiple-passages" prototype strains have not biased our perspective of the *in vivo* evolutionary pathway behind the *C. trachomatis* diversification. While this emerged as an important conclusion, the results from the Chapter VIII further pointed out that the *C. trachomatis* transition from the *in vivo* to the *in vitro* environment has the potential to largely affect the phenotype, potentially yielding virulence attenuation. In fact, we observed a tropism-specific rampant inactivation of the virulence gene CT135, with CT135-negative clones emerging for all epithelial-genital populations (but not for LGV and ocular populations). Additional global transcriptomic analyses revealed that CT135 deletion impacted the expression of multiple

virulence genes essentially associated with the invasion of the host-cell. We also observed that chromosomal genes known to be regulated by the plasmid were affected, suggesting that the plasmid-mediated virulence could also be smoothed in the *C. trachomatis in vivo* to *in vitro* transition. On one hand, these results potentially disclosed the molecular basis underlying the CT135-mediated virulence, opening new questions whether this uncharacterized gene plays a role in defining the tissue-tropism of *C. trachomatis*. On the other hand, this scenario raises an issue of special relevance considering the recent progress in *Chlamydia* mutagenesis and transformation, since the selection of mutants/transformants demands culture propagation. Hence, in order to circumvent the selection of CT135-negative mutants, likely yielding less virulent strains, it would be prudent to restrict culture propagation to minimal passages and to verify the status of the CT135 genotype. These approaches might be beneficial for interpreting phenotypic data and cross-compare results between different laboratories. Another important result of the Chapter VIII concerns the hypothesis that the functionality of two proteins potentially involved in pathogenesis, the cytotoxin and a lipid A biosynthetic enzyme, is regulated by phase variation mechanisms. This brings to discussion whether phase-variation mechanisms could modulate important *C. trachomatis* functions by yielding variable phenotypes with impact on bacterial adaptation and pathogenesis. On the other hand, it is illustrative that advances in genome sequencing technologies have largely increased our power to raise new relevant questions to be addressed in future follow-up investigations aiming to clarify the mechanisms by which *C. trachomatis* cause human disease.

In conclusion, the findings presented in this Ph.D. thesis may contribute for the understanding of *C. trachomatis* adaptive evolution and provide new insights into the biological role of *C. trachomatis* hypothetical proteins. The perfect knowledge of the determinants of tissue tropism, and of virulence, pathogenic or ecological success dissimilarities among *C. trachomatis* strains remains to be understood, but certainly it should involve the bacterial players highlighted throughout this Ph.D. dissertation. On the other hand, the findings obtained on behalf of the study of the *in vitro* adaptation may immediately impact our view regarding the *C. trachomatis* population dynamics and help us to better interpret results from future *in vitro* or *in vivo* studies.

Future directions

Our ability to acquire knowledge about the *C. trachomatis* pathobiology has now entered in a new level due to the rampant progress in genetic manipulation and mutagenesis technologies, which are expected to become easily available to the scientific community in the next years. Nevertheless, it is likely that some historical enigmas will remain for a while, such as: *i*) Why strains from LGV serovars are able to infect the macrophages but not the strains from ocular and genital serovars?; *ii*) Why LGV strains are not detected in ocular infections? Are they unable to infect the eye?; *iii*) Why ocular infections with strains from genital serovars do not progress to trachoma?; and *iv*) Which bacterial factors determine the high clinical prevalence of serovars E and F in the mucosa-restricted genital STIs, and L2b strains in the MSM-associated rectal STIs and typical LGV cases?. In turn, some

questions raised on the course of this Ph.D. thesis, namely those arising from the investigation of *C. trachomatis in vivo* to *in vitro* transition (Chapter VIII), will soon constitute starting points for research lines of our group. In particular, we have two major goals for the near future:

- To understand the role of CT135 in virulence and tissue tropism:

Approaches:

- Performing extent analyses of laboratory adaptation using cell lines representing human tissue other than the genitalia. Questions: Is indeed CT135 a determinant of tissue tropism? Is the tropism-specific loss of the virulence gene CT135 extensive to other human cell lines, such human monocytic or human conjunctival cell lines?

- Performing *in vivo* studies. Questions: How much advantage does CT135 confer for *in vivo* growth of *C. trachomatis*? In an *in vivo* inoculation with a mixture of CT135-negative and CT135-positive clones, which clones will proliferate?

- To understand the role of phase variation in *C. trachomatis* adaptation and pathogenesis

Approaches:

- Surveying the cytotoxin "genotype" in multiple clinical samples. Questions: Which is the proportion of *in vivo* populations with variability affecting the cytotoxin? Is there any correlation between the predicted "ON" or "OFF" phenotype with particular clinical manifestations?

- **Studying the potential role of the cytotoxin in the adaptation to tryptophan availability.** Question: Will we observe a switch from the "OFF" to the "ON" status of the cytotoxin upon tryptophan depletion through the action of IFN-γ?

- Studying the impact of phase variation on the LPS phenotype. Questions: Is the expression of gene *lpxC* affected by the variable homopolymeric tract in its promoter region? Is the immunoreactivity of LPS affected?

Final author's note:

Although most of the post-doctoral period will be dedicated to the use of bioinformatics to investigate other pathogens with impact on public health, the above cited research lines in *C. trachomatis* will constitute an important parallel research activity involving both the application to funding programs and supervision of students.

References

References

1. **Sachse, K., K. Laroucau, K. Riege, S. Wehner, M. Dilcher, H. H. Creasy, M. Weidmann, G. Myers, F. Vorimore, N. Vicari, S. Magnino, E. Liebler-Tenorio, A. Ruettger, P. M. Bavoil, F. T. Hufert, R. Rossello-Mora, and M. Marz**. 2014. 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. **37**:79-88.
2. **Sachse, K., P. M. Bavoil, B. Kaltenboeck, R. S. Stephens, C. C. Kuo, R. Rossello-Mora, and M. Horn**. 2015. Emendation of the family *Chlamydiaceae*: proposal of a single genus, *Chlamydia*, to include all currently recognized species. Syst. Appl. Microbiol. **38**:99-103.
3. **Stephens, R. S., G. Myers, M. Eppinger, and P. M. Bavoil**. 2009. Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved. FEMS Immunol. Med. Microbiol. **55**:115-119.
4. **Bebear, C. and B. B. de**. 2009. Genital *Chlamydia trachomatis* infections. Clin. Microbiol. Infect. **15**:4-10.
5. **White, J. A.** 2009. Manifestations and management of lymphogranuloma venereum. Curr. Opin. Infect. Dis. **22**:57-66.
6. **Wright, H. R., A. Turner, and H. R. Taylor**. 2008. Trachoma. Lancet **371**:1945-1954.
7. **Campbell, L. A. and M. E. Rosenfeld**. 2014. Persistent C. *pneumoniae* infection in atherosclerotic lesions: rethinking the clinical trials. Front Cell Infect. Microbiol. **4**:34.
8. **Rodriguez, A. R., G. Plascencia-Villa, C. M. Witt, J. J. Yu, M. Jose-Yacaman, J. P. Chambers, G. Perry, M. N. Guentzel, and B. P. Arulanandam**. 2015. *Chlamydia pneumoniae* promotes dysfunction of pancreatic beta cells. Cell Immunol. **295**:83-91.
9. **Saikku, P.** 1999. Epidemiology of *Chlamydia pneumoniae* in atherosclerosis. Am. Heart J. **138**:S500-S503.
10. **Sethi, S. and T. F. Murphy**. 2001. Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review. Clin. Microbiol. Rev. **14**:336-363.
11. **Nunes, A. and J. P. Gomes**. 2014. Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. Infect. Genet. Evol. **23**:49-64.
12. **Petrovay, F. and E. Balla**. 2008. Two fatal cases of psittacosis caused by *Chlamydophila psittaci*. J. Med. Microbiol. **57**:1296-1298.
13. **Van, D. C., D. S. Beeckman, K. Verminnen, M. Marien, H. Nauwynck, L. T. Boesinghe, and D. Vanrompay**. 2009. Simultaneous zoonotic transmission of *Chlamydophila psittaci* genotypes D, F and E/B to a veterinary scientist. Vet. Microbiol. **135**:78-81.
14. **Vorimore, F., A. Thebault, S. Poisson, D. Cleva, J. Robineau, B. B. de, B. Durand, and K. Laroucau**. 2015. *Chlamydia psittaci* in ducks: a hidden health risk for poultry workers. Pathog. Dis. **73**:1-9.
15. **Voigt, A., G. Schofl, and H. P. Saluz**. 2012. The *Chlamydia psittaci* genome: a comparative analysis of intracellular pathogens. PLoS. One. **7**:e35097.
16. **Aitken ID and Longbottom D**. 2007. Chlamydial abortion, p. 105-111. *In* Aitken ID (ed.). Oxford: Blackwell Publishing.
17. **Kerr, K., G. Entrican, D. McKeever, and D. Longbottom**. 2005. Immunopathology of *Chlamydophila* abortus infection in sheep and mice. Res. Vet. Sci. **78**:1-7.
18. **Longbottom, D. and L. J. Coulter**. 2003. Animal chlamydioses and zoonotic implications. J. Comp Pathol. **128**:217-244.
19. **Browning, G. F.** 2004. Is *Chlamydophila* felis a significant zoonotic pathogen? Aust. Vet. J. **82**:695-696.
20. **Sykes, J. E.** 2005. Feline chlamydiosis. Clin. Tech. Small Anim Pract. **20**:129-134.
21. **Wu, S. M., S. Y. Huang, M. J. Xu, D. H. Zhou, H. Q. Song, and X. Q. Zhu**. 2013. *Chlamydia felis* exposure in companion dogs and cats in Lanzhou, China: a public health concern. BMC. Vet. Res. **9**:104.
22. **Bachmann, N. L., T. A. Fraser, C. Bertelli, M. Jelocnik, A. Gillett, O. Funnell, C. Flanagan, G. S. Myers, P. Timms, and A. Polkinghorne**. 2014. Comparative genomics of koala, cattle and sheep strains of *Chlamydia* pecorum. BMC. Genomics **15**:667.
23. **Mojica, S., C. H. Huot, S. Daugherty, T. D. Read, T. Kim, B. Kaltenboeck, P. Bavoil, and G. S. Myers**. 2011. Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. J. Bacteriol. **193**:3690.
24. **Storz J**. 1988. Overview of animal diseases induced by chlamydial infections, p. 167-192. *In* Barron A.I. (ed.), Microbiology of *Chlamydia*. CRC Press, Boca Raton, Florida.

25. **Rogers, D. G., A. A. Andersen, and B. D. Hunsaker**. 1996. Lung and nasal lesions caused by a swine chlamydial isolate in gnotobiotic pigs. J. Vet. Diagn. Invest **8**:45-55.
26. **Schautteet, K. and D. Vanrompay**. 2011. *Chlamydia*ceae infections in pig. Vet. Res. **42**:29.
27. **Sachse, K. and K. Laroucau**. 2015. Two more species of *Chlamydia*-does it make a difference? Pathog. Dis. **73**:1-3.
28. **Read, T. D., G. S. Myers, R. C. Brunham, W. C. Nelson, I. T. Paulsen, J. Heidelberg, E. Holtzapple, H. Khouri, N. B. Federova, H. A. Carty, L. A. Umayam, D. H. Haft, J. Peterson, M. J. Beanan, O. White, S. L. Salzberg, R. C. Hsia, G. McClarty, R. G. Rank, P. M. Bavoil, and C. M. Fraser**. 2003. Genome sequence of *Chlamydophila* caviae (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the *Chlamydia*ceae. Nucleic Acids Res. **31**:2134-2147.
29. **Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser**. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. **28**:1397-1406.
30. **Barron, A. L., H. J. White, R. G. Rank, B. L. Soloff, and E. B. Moses**. 1981. A new animal model for the study of *Chlamydia trachomatis* genital infections: infection of mice with the agent of mouse pneumonitis. J. Infect. Dis. **143**:63-66.
31. **Morrison, R. P. and H. D. Caldwell**. 2002. Immunity to murine chlamydial genital infection. Infect. Immun. **70**:2741-2751.
32. **Ramsey, K. H., I. M. Sigar, J. H. Schripsema, C. J. Denman, A. K. Bowlin, G. A. Myers, and R. G. Rank**. 2009. Strain and virulence diversity in the mouse pathogen *Chlamydia muridarum*. Infect. Immun. **77**:3284-3293.
33. **De, C. E., I. Kalmar, and D. Vanrompay**. 2013. Animal models for studying female genital tract infection with *Chlamydia trachomatis*. Infect. Immun. **81**:3060-3067.
34. **Yuan, Y., Y. X. Zhang, N. G. Watkins, and H. D. Caldwell**. 1989. 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. **57**:1040-1049.
35. **Brunelle, B. W. and G. F. Sensabaugh**. 2006. The *ompA* gene in *Chlamydia trachomatis* differs in phylogeny and rate of evolution from other regions of the genome. Infect. Immun. **74**:578-585.
36. **Fitch, W. M., E. M. Peterson, and L. M. de la Maza**. 1993. Phylogenetic analysis of the outer-membrane-protein genes of *Chlamydia*e, and its implication for vaccine development. Mol. Biol. Evol. **10**:892-913.
37. **Harris, S. R., I. N. Clarke, H. M. Seth-Smith, A. W. Solomon, L. T. Cutcliffe, P. Marsh, R. J. Skilton, M. J. Holland, D. Mabey, R. W. Peeling, D. A. Lewis, B. G. Spratt, M. Unemo, K. Persson, C. Bjartling, R. Brunham, H. J. de Vries, S. A. Morre, A. Speksnijder, C. M. Bebear, M. Clerc, B. B. de, J. Parkhill, and N. R. Thomson**. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nat. Genet. **44**:413-9, S1.
38. **Millman, K. L., S. Tavare, and D. Dean**. 2001. Recombination in the *ompA* gene but not the omcB gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. J. Bacteriol. **183**:5997-6008.
39. **Nunes, A., M. J. Borrego, B. Nunes, C. Florindo, and J. P. Gomes**. 2009. Evolutionary dynamics of *ompA*, the gene encoding the *Chlamydia trachomatis* key antigen. J. Bacteriol. **191**:7182-7192.
40. **Stothard, D. R., G. Boguslawski, and R. B. Jones**. 1998. Phylogenetic analysis of the *Chlamydia trachomatis* major outer membrane protein and examination of potential pathogenic determinants. Infect. Immun. **66**:3618-3625.
41. **Burton, M. J.** 2007. Trachoma: an overview. Br. Med. Bull. **84**:99-116.
42. **Mariotti, S. P., D. Pascolini, and J. Rose-Nussbaumer**. 2009. Trachoma: global magnitude of a preventable cause of blindness. Br. J. Ophthalmol. **93**:563-568.
43. **Pascolini, D. and S. P. Mariotti**. 2012. Global estimates of visual impairment: 2010. Br. J. Ophthalmol. **96**:614-618.
44. **World Health Organization**. 2011. Prevalence and Incidence of Selected Sexually Transmitted Infections. (World Health Organization, Geneva, Switzerland, 2011) .
45. **Mabey, D. and R. W. Peeling**. 2002. Lymphogranuloma venereum. Sex Transm. Infect. **78**:90-92.

46. **Thomson, N. R., M. T. Holden, C. Carder, N. Lennard, S. J. Lockey, P. Marsh, P. Skipp, C. D. O'Connor, I. Goodhead, H. Norbertzcak, B. Harris, D. Ormond, R. Rance, M. A. Quail, J. Parkhill, R. S. Stephens, and I. N. Clarke**. 2008. *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res. **18**:161-171.
47. **Schachter, J.** 1978. Chlamydial infections (first of three parts). N. Engl. J. Med. **298**:428-435.
48. **Nieuwenhuis, R. F., J. M. Ossewaarde, W. I. van der Meijden, and H. A. Neumann**. 2003. Unusual presentation of early lymphogranuloma venereum in an HIV-1 infected patient: effective treatment with 1 g azithromycin. Sex Transm. Infect. **79**:453-455.
49. **Dal, C., I, M. Mistrangelo, C. Cariti, M. Chiriotto, A. Lucchini, M. Vigna, M. Morino, and P. G. Di**. 2014. Lymphogranuloma venereum: an old, forgotten re-emerging systemic disease. Panminerva Med. **56**:73-83.
50. **Schachter J**. 1999. Infection and disease epidemiology, p. 139-169. *In* R. S. Stephens (ed.), *Chlamydia*: intracellular biology, pathogenesis and immunity. ASM Press, Washington, DC.
51. **Caldwell, H. D., H. Wood, D. Crane, R. Bailey, R. B. Jones, D. Mabey, I. Maclean, Z. Mohammed, R. Peeling, C. Roshick, J. Schachter, A. W. Solomon, W. E. Stamm, R. J. Suchland, L. Taylor, S. K. West, T. C. Quinn, R. J. Belland, and G. McClarty**. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J. Clin. Invest **111**:1757-1769.
52. **Porter, M., D. Mak, G. Chidlow, G. B. Harnett, and D. W. Smith**. 2008. The molecular epidemiology of ocular *Chlamydia trachomatis* infections in Western Australia: implications for trachoma control. Am. J. Trop. Med. Hyg. **78**:514-517.
53. **Andreasen, A. A., M. J. Burton, M. J. Holland, S. Polley, N. Faal, D. C. Mabey, and R. L. Bailey**. 2008. *Chlamydia trachomatis ompA* variants in trachoma: what do they tell us? PLoS. Negl. Trop. Dis. **2**:e306.
54. **World Health Organization**. 2012. Accelerating work to overcome the global impact of neglected tropical diseases - A roadmap for implementation. (World Health Organization, Geneva, Switzerland, 2012) .
55. **World Health Organization**. 1998. Global elimination of blinding trachoma. Resolution WHA 51.11 adopted by World Health Assembly. 16 May 1998. (World Health Organization, Geneva, Switzerland, 1998) .
56. **van de Laar, M. J. and S. A. Morre**. 2007. *Chlamydia*: a major challenge for public health. Euro. Surveill **12**:E1-E2.
57. **Peipert, J. F.** 2003. Clinical practice. Genital chlamydial infections. N. Engl. J. Med. **349**:2424-2430.
58. **Shaw, K., D. Coleman, M. O'Sullivan, and N. Stephens**. 2011. Public health policies and management strategies for genital *Chlamydia trachomatis* infection. Risk Manag. Healthc. Policy **4**:57-65.
59. **Taylor-Robinson, D. and B. J. Thomas**. 1980. The role of *Chlamydia trachomatis* in genital-tract and associated diseases. J. Clin. Pathol. **33**:205-233.
60. **Christerson, L., R. J. Bom, S. M. Bruisten, R. Yass, J. Hardick, G. Bratt, C. A. Gaydos, S. A. Morre, and B. Herrmann**. 2012. *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. **50**:3548-3555.
61. **Lysen, M., A. Osterlund, C. J. Rubin, T. Persson, I. Persson, and B. Herrmann**. 2004. Characterization of *ompA* genotypes by sequence analysis of DNA from all detected cases of *Chlamydia trachomatis* infections during 1 year of contact tracing in a Swedish County. J. Clin. Microbiol. **42**:1641-1647.
62. **Morre, S. A., L. Rozendaal, V. van, I, A. J. Boeke, V. van, V, J. Schirm, B. S. de, J. A. van Den Hoek, G. J. van Doornum, C. J. Meijer, and A. J. van den Brule**. 2000. Urogenital *Chlamydia trachomatis* serovars in men and women with a symptomatic or asymptomatic infection: an association with clinical manifestations? J. Clin. Microbiol. **38**:2292-2296.
63. **Nunes, A., P. J. Nogueira, M. J. Borrego, and J. P. Gomes**. 2010. Adaptive evolution of the *Chlamydia trachomatis* dominant antigen reveals distinct evolutionary scenarios for B- and T-cell epitopes: worldwide survey. PLoS. One. **5**.
64. **Quint, K. D., R. J. Bom, W. G. Quint, S. M. Bruisten, M. F. van der Loeff, S. A. Morre, and H. J. de Vries**. 2011. Anal infections with concomitant *Chlamydia trachomatis*

genotypes among men who have sex with men in Amsterdam, the Netherlands. BMC. Infect. Dis. **11**:63.
65. **Yang, B., H. P. Zheng, Z. Q. Feng, Y. H. Xue, X. Z. Wu, J. M. Huang, X. J. Xue, and H. N. Jiang**. 2010. The prevalence and distribution of *Chlamydia trachomatis* genotypes among sexually transmitted disease clinic patients in Guangzhou, China, 2005-2008. Jpn. J. Infect. Dis. **63**:342-345.
66. **de Vries, H. J., A. Zingoni, A. Kreuter, H. Moi, and J. A. White**. 2015. 2013 European guideline on the management of lymphogranuloma venereum. J. Eur. Acad. Dermatol. Venereol. **29**:1-6.
67. **Savage, E. J., M. J. van de Laar, A. Gallay, S. M. van der, O. Hamouda, A. Sasse, S. Hoffmann, M. Diez, M. J. Borrego, C. M. Lowndes, and C. Ison**. 2009. Lymphogranuloma venereum in Europe, 2003-2008. Euro. Surveill **14**.
68. **de, L. M. and J. Nougue**. 2013. [Lymphogranuloma venereum: new serovariant L2b and old "groove sign"]. Bull. Soc. Pathol. Exot. **106**:153-155.
69. **Marcotte, T., Y. Lee, M. Pandori, V. Jain, and S. E. Cohen**. 2014. Lymphogranuloma venereum causing a persistent genital ulcer. Sex Transm. Dis. **41**:280-282.
70. **Peuchant, O., C. Baldit, R. C. Le, S. Trombert-Paolantoni, M. Clerc, C. Bebear, and B. B. de**. 2011. First case of *Chlamydia trachomatis* L2b proctitis in a woman. Clin. Microbiol. Infect. **17**:E21-E23.
71. **Verweij, S. P., S. Ouburg, V. H. de, S. A. Morre, C. J. van Ginkel, H. Bos, and F. W. Sebens**. 2012. The first case record of a female patient with bubonic lymphogranuloma venereum (LGV), serovariant L2b. Sex Transm. Infect. **88**:346-347.
72. **Gomes, J. P., A. Nunes, C. Florindo, M. A. Ferreira, I. Santo, J. Azevedo, and M. J. Borrego**. 2009. Lymphogranuloma venereum in Portugal: unusual events and new variants during 2007. Sex Transm. Dis. **36**:88-91.
73. **Mechai, F., B. B. de, O. Aoun, A. Merens, P. Imbert, and C. Rapp**. 2010. Doxycycline failure in lymphogranuloma venereum. Sex Transm. Infect. **86**:278-279.
74. **Oud, E. V., N. H. de Vrieze, M. A. de, and H. J. de Vries**. 2014. Pitfalls in the diagnosis and management of inguinal lymphogranuloma venereum: important lessons from a case series. Sex Transm. Infect. **90**:279-282.
75. **Carter, J. D. and R. D. Inman**. 2011. *Chlamydia*-induced reactive arthritis: hidden in plain sight? Best. Pract. Res. Clin. Rheumatol. **25**:359-374.
76. **Gerard, H. C., J. A. Stanich, J. A. Whittum-Hudson, H. R. Schumacher, J. D. Carter, and A. P. Hudson**. 2010. Patients with *Chlamydia*-associated arthritis have ocular (trachoma), not genital, serovars of C. *trachomatis* in synovial tissue. Microb. Pathog. **48**:62-68.
77. **Pendle, S. and A. Gowers**. 2012. Reactive arthritis associated with proctitis due to *Chlamydia trachomatis* serovar L2b. Sex Transm. Dis. **39**:79-80.
78. **Taylor-Robinson, D. and A. Keat**. 2015. Observations on *Chlamydia trachomatis* and other microbes in reactive arthritis. Int. J. STD AIDS **26**:139-144.
79. **Laga, M., N. Nzila, and J. Goeman**. 1991. The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. AIDS **5 Suppl 1**:S55-S63.
80. **Baeten, J. M. and J. Overbaugh**. 2003. Measuring the infectiousness of persons with HIV-1: opportunities for preventing sexual HIV-1 transmission. Curr. HIV. Res. **1**:69-86.
81. **Xiridou, M., H. J. Vriend, A. K. Lugner, J. Wallinga, J. S. Fennema, J. M. Prins, S. E. Geerlings, B. J. Rijnders, M. Prins, H. J. de Vries, M. J. Postma, M. G. van Veen, Schim van der Loeff MF, and M. A. van der Sande**. 2013. Modelling the impact of *Chlamydia* screening on the transmission of HIV among men who have sex with men. BMC. Infect. Dis. **13**:436.
82. **Luostarinen, T., P. B. Namujju, M. Merikukka, J. Dillner, T. Hakulinen, P. Koskela, J. Paavonen, H. M. Surcel, and M. Lehtinen**. 2013. Order of HPV/*Chlamydia* infections and cervical high-grade precancer risk: a case-cohort study. Int. J. Cancer **133**:1756-1759.
83. **Silva, J., F. Cerqueira, and R. Medeiros**. 2014. *Chlamydia trachomatis* infection: implications for HPV status and cervical cancer. Arch. Gynecol. Obstet. **289**:715-723.
84. **Chumduri, C., R. K. Gurumurthy, P. K. Zadora, Y. Mi, and T. F. Meyer**. 2013. *Chlamydia* infection promotes host DNA damage and proliferation but impairs the DNA damage response. Cell Host. Microbe **13**:746-758.

85. **Gonzalez, E., M. Rother, M. C. Kerr, M. A. Al-Zeer, M. bu-Lubad, M. Kessler, V. Brinkmann, A. Loewer, and T. F. Meyer**. 2014. *Chlamydia* infection depends on a functional MDM2-p53 axis. Nat. Commun. **5**:5201.
86. **Shanmughapriya, S., G. Senthilkumar, K. Vinodhini, B. C. Das, N. Vasanthi, and K. Natarajaseenivasan**. 2012. Viral and bacterial aetiologies of epithelial ovarian cancer. Eur. J. Clin. Microbiol. Infect. Dis. **31**:2311-2317.
87. **Liechti, G. W., E. Kuru, E. Hall, A. Kalinda, Y. V. Brun, M. VanNieuwenhze, and A. T. Maurelli**. 2014. A new metabolic cell-wall labelling method reveals peptidoglycan in *Chlamydia trachomatis*. Nature **506**:507-510.
88. **AbdelRahman, Y. M. and R. J. Belland**. 2005. The chlamydial developmental cycle. FEMS Microbiol. Rev. **29**:949-959.
89. **Schachter, J.** 1988. The intracellular life of *Chlamydia*. Curr. Top. Microbiol. Immunol. **138**:109-139.
90. **Bedson SP and Bland JOW**. 1932. A morphological study of psittacosis virus, with the description of a developmental cycle. Br. J. Exp. Pathol. 461-466.
91. **Moulder, J. W.** 1991. Interaction of *Chlamydia*e and host cells *in vitro*. Microbiol. Rev. **55**:143-190.
92. **Barry, C. E., III, S. F. Hayes, and T. Hackstadt**. 1992. Nucleoid condensation in *Escherichia coli* that express a chlamydial histone homolog. Science **256**:377-379.
93. **Hackstadt, T., W. J. Todd, and H. D. Caldwell**. 1985. Disulfide-mediated interactions of the *Chlamydia*l major outer membrane protein: role in the differentiation of *Chlamydiae*? J. Bacteriol. **161**:25-31.
94. **Omsland, A., J. Sager, V. Nair, D. E. Sturdevant, and T. Hackstadt**. 2012. Developmental stage-specific metabolic and transcriptional activity of *Chlamydia trachomatis* in an axenic medium. Proc. Natl. Acad. Sci. U. S. A **109**:19781-19785.
95. **Hybiske, K. and R. S. Stephens**. 2007. Mechanisms of host cell exit by the intracellular bacterium *Chlamydia*. Proc. Natl. Acad. Sci. U. S. A **104**:11430-11435.
96. **Hogan, R. J., S. A. Mathews, S. Mukhopadhyay, J. T. Summersgill, and P. Timms**. 2004. Chlamydial persistence: beyond the biphasic paradigm. Infect. Immun. **72**:1843-1855.
97. **Wyrick, P. B.** 2010. *Chlamydia trachomatis* persistence *in vitro*: an overview. J. Infect. Dis. **201 Suppl 2**:S88-S95.
98. **Shaw, E. I., C. A. Dooley, E. R. Fischer, M. A. Scidmore, K. A. Fields, and T. Hackstadt**. 2000. Three temporal classes of gene expression during the *Chlamydia trachomatis* developmental cycle. Mol. Microbiol. **37**:913-925.
99. **Belland, R. J., G. Zhong, D. D. Crane, D. Hogan, D. Sturdevant, J. Sharma, W. L. Beatty, and H. D. Caldwell**. 2003. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S. A **100**:8478-8483.
100. **Nicholson, T. L., L. Olinger, K. Chong, G. Schoolnik, and R. S. Stephens**. 2003. Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. J. Bacteriol. **185**:3179-3189.
101. **Humphrys, M. S., T. Creasy, Y. Sun, A. C. Shetty, M. C. Chibucos, E. F. Drabek, C. M. Fraser, U. Farooq, N. Sengamalay, S. Ott, H. Shou, P. M. Bavoil, A. Mahurkar, and G. S. Myers**. 2013. Simultaneous transcriptional profiling of bacteria and their host cells. PLoS. One. **8**:e80597.
102. **Albrecht, M., C. M. Sharma, R. Reinhardt, J. Vogel, and T. Rudel**. 2010. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res. **38**:868-877.
103. **Valdivia, R. H.** 2008. *Chlamydia* effector proteins and new insights into chlamydial cellular microbiology. Curr. Opin. Microbiol. **11**:53-59.
104. **Mathews S and Timms P**. 2006. *In silico* identification of chlamydial promoters and their role in regulation of development, *In* Bavoil PM and Wyrick PB (ed.), *Chlamydia*: genomics and pathogenesis. Horizon Bioscience, Norfolk, United Kingdom.
105. **Tan**. 2006. Regulation of gene expression, *In* Bavoil PM and W. P. (ed.), *Chlamydia*: Genomics and Pathogenesis. Horizon Scientific Press.
106. **Tan M**. 2012. Temporal gene regulation during the chlamydial development cycle., p. 149-169. *In* Tan M and Bavoil PM (ed.), Intracellular Pathogens I: *Chlamydiales*. ASM Press, Washington, D.C.
107. **Fadel, S. and A. Eley**. 2007. *Chlamydia trachomatis* OmcB protein is a surface-exposed glycosaminoglycan-dependent adhesin. J. Med. Microbiol. **56**:15-22.

108. **Fechtner, T., S. Stallmann, K. Moelleken, K. L. Meyer, and J. H. Hegemann**. 2013. Characterization of the interaction between the chlamydial adhesin OmcB and the human host cell. J. Bacteriol. **195**:5323-5333.
109. **Becker, E. and J. H. Hegemann**. 2014. All subtypes of the Pmp adhesin family are implicated in chlamydial virulence and show species-specific function. Microbiologyopen. **3**:544-556.
110. **Crane, D. D., J. H. Carlson, E. R. Fischer, P. Bavoil, R. C. Hsia, C. Tan, C. C. Kuo, and H. D. Caldwell**. 2006. *Chlamydia trachomatis* polymorphic membrane protein D is a species-common pan-neutralizing antigen. Proc. Natl. Acad. Sci. U. S. A **103**:1894-1899.
111. **Gomes, J. P., A. Nunes, W. J. Bruno, M. J. Borrego, C. Florindo, and D. Dean**. 2006. 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. **188**:275-286.
112. **Grimwood, J. and R. S. Stephens**. 1999. Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. Microb. Comp Genomics **4**:187-201.
113. **Barry, C. E., III, T. J. Brickman, and T. Hackstadt**. 1993. Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. Mol. Microbiol. **9**:273-283.
114. **Brickman, T. J., C. E. Barry, III, and T. Hackstadt**. 1993. Molecular cloning and expression of *hctB* encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. J. Bacteriol. **175**:4274-4281.
115. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis**. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282**:754-759.
116. **Kubo, A. and R. S. Stephens**. 2000. Characterization and functional analysis of PorB, a *Chlamydia* porin and neutralizing target. Mol. Microbiol. **38**:772-780.
117. **Grieshaber, N. A., S. S. Grieshaber, E. R. Fischer, and T. Hackstadt**. 2006. A small RNA inhibits translation of the histone-like protein Hc1 in *Chlamydia trachomatis*. Mol. Microbiol. **59**:541-550.
118. **Tattersall, J., G. V. Rao, J. Runac, T. Hackstadt, S. S. Grieshaber, and N. A. Grieshaber**. 2012. Translation inhibition of the developmental cycle protein HctA by the small RNA IhtA is conserved across *Chlamydia*. PLoS. One. **7**:e47439.
119. **Rosario, C. J., B. R. Hanson, and M. Tan**. 2014. The transcriptional repressor EUO regulates both subsets of *Chlamydia* late genes. Mol. Microbiol. **94**:888-897.
120. **Rosario, C. J. and M. Tan**. 2012. The early gene product EUO is a transcriptional repressor that selectively regulates promoters of *Chlamydia* late genes. Mol. Microbiol. **84**:1097-1107.
121. **Bastidas, R. J., C. A. Elwell, J. N. Engel, and R. H. Valdivia**. 2013. Chlamydial intracellular survival strategies. Cold Spring Harb. Perspect. Med. **3**:a010256.
122. **Cocchiaro, J. L. and R. H. Valdivia**. 2009. New insights into *Chlamydia* intracellular survival mechanisms. Cell Microbiol. **11**:1571-1578.
123. **Derre, I.** 2015. *Chlamydiae* interaction with the Endoplasmic Reticulum: contact, function and consequences. Cell Microbiol.
124. **Mehlitz, A. and T. Rudel**. 2013. Modulation of host signaling and cellular responses by *Chlamydia*. Cell Commun. Signal. **11**:90.
125. **Saka, H. A. and R. H. Valdivia**. 2010. Acquisition of nutrients by *Chlamydiae*: unique challenges of living in an intracellular compartment. Curr. Opin. Microbiol. **13**:4-10.
126. **Scidmore, M. A.** 2011. Recent advances in *Chlamydia* subversion of host cytoskeletal and membrane trafficking pathways. Microbes. Infect. **13**:527-535.
127. **Betts, H. J., K. Wolf, and K. A. Fields**. 2009. Effector protein modulation of host cells: examples in the *Chlamydia* spp. arsenal. Curr. Opin. Microbiol. **12**:81-87.
128. **Moore, E. R. and S. P. Ouellette**. 2014. Reconceptualizing the chlamydial inclusion as a pathogen-specified parasitic organelle: an expanded role for Inc proteins. Front Cell Infect. Microbiol. **4**:157.
129. **Dehoux, P., R. Flores, C. Dauga, G. Zhong, and A. Subtil**. 2011. Multi-genome identification and characterization of *Chlamydiae*-specific type III secretion substrates: the Inc proteins. BMC. Genomics **12**:109.
130. **Fields, K. A., D. J. Mead, C. A. Dooley, and T. Hackstadt**. 2003. *Chlamydia trachomatis* type III secretion: evidence for a functional apparatus during early-cycle development. Mol. Microbiol. **48**:671-683.

131. **Rockey, D. D., M. A. Scidmore, J. P. Bannantine, and W. J. Brown**. 2002. Proteins in the chlamydial inclusion membrane. Microbes. Infect. **4**:333-340.
132. **Subtil, A., C. Delevoye, M. E. Balana, L. Tastevin, S. Perrinet, and A. utry-Varsat**. 2005. A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. Mol. Microbiol. **56**:1636-1647.
133. **Heuer, D., V. Brinkmann, T. F. Meyer, and A. J. Szczepek**. 2003. Expression and translocation of chlamydial protease during acute and persistent infection of the epithelial HEp-2 cells with *Chlamydophila* (*Chlamydia*) *pneumoniae*. Cell Microbiol. **5**:315-322.
134. **Lad, S. P., G. Yang, D. A. Scott, G. Wang, P. Nair, J. Mathison, V. S. Reddy, and E. Li**. 2007. chlamydial CT441 is a PDZ domain-containing tail-specific protease that interferes with the NF-kappaB pathway of immune response. J. Bacteriol. **189**:6619-6625.
135. **Zhong, G.** 2011. *Chlamydia trachomatis* secretion of proteases for manipulating host signaling pathways. Front Microbiol. **2**:14.
136. **Snavely, E. A., M. Kokes, J. D. Dunn, H. A. Saka, B. D. Nguyen, R. J. Bastidas, D. G. McCafferty, and R. H. Valdivia**. 2014. Reassessing the role of the secreted protease CPAF in *Chlamydia trachomatis* infection through genetic approaches. Pathog. Dis. **71**:336-351.
137. **Mueller, K. E., G. V. Plano, and K. A. Fields**. 2014. New frontiers in type III secretion biology: the *Chlamydia* perspective. Infect. Immun. **82**:2-9.
138. **Peters, J., D. P. Wilson, G. Myers, P. Timms, and P. M. Bavoil**. 2007. Type III secretion a la *Chlamydia*. Trends Microbiol. **15**:241-251.
139. **Cornelis, G. R.** 2006. The type III secretion injectisome. Nat. Rev. Microbiol. **4**:811-825.
140. **Clifton, D. R., K. A. Fields, S. S. Grieshaber, C. A. Dooley, E. R. Fischer, D. J. Mead, R. A. Carabeo, and T. Hackstadt**. 2004. A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proc. Natl. Acad. Sci. U. S. A **101**:10166-10171.
141. **Jewett, T. J., N. J. Miller, C. A. Dooley, and T. Hackstadt**. 2010. The conserved Tarp actin binding domain is important for chlamydial invasion. PLoS. Pathog. **6**:e1000997.
142. **Hackstadt, T., M. A. Scidmore-Carlson, E. I. Shaw, and E. R. Fischer**. 1999. The *Chlamydia trachomatis* IncA protein is required for homotypic vesicle fusion. Cell Microbiol. **1**:119-130.
143. **Ronzone, E., J. Wesolowski, L. D. Bauler, A. Bhardwaj, T. Hackstadt, and F. Paumet**. 2014. An alpha-helical core encodes the dual functions of the chlamydial protein IncA. J. Biol. Chem. **289**:33469-33480.
144. **Ronzone, E. and F. Paumet**. 2013. Two coiled-coil domains of *Chlamydia trachomatis* IncA affect membrane fusion events during infection. PLoS. One. **8**:e69769.
145. **Suchland, R. J., D. D. Rockey, J. P. Bannantine, and W. E. Stamm**. 2000. Isolates of *Chlamydia trachomatis* that occupy nonfusogenic inclusions lack IncA, a protein localized to the inclusion membrane. Infect. Immun. **68**:360-367.
146. **Abdelsamed, H., J. Peters, and G. I. Byrne**. 2013. Genetic variation in *Chlamydia trachomatis* and their hosts: impact on disease severity and tissue tropism. Future. Microbiol. **8**:1129-1146.
147. **Asner, S. A., S. A. Morre, P. Y. Bochud, and G. Greub**. 2014. Host factors and genetic susceptibility to infections due to intracellular bacteria and fastidious organisms. Clin. Microbiol. Infect. **20**:1246-1253.
148. **Andersson, J. O. and S. G. Andersson**. 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9**:664-671.
149. **Zomorodipour, A. and S. G. Andersson**. 1999. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. FEBS Lett. **452**:11-15.
150. **Felsenstein, J.** 1974. The evolutionary advantage of recombination. Genetics **78**:737-756.
151. **MULLER, H. J.** 1964. THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. Mutat. Res. **106**:2-9.
152. **Moran, N. A.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A **93**:2873-2878.
153. **Andersson, D. I. and D. Hughes**. 1996. Muller's ratchet decreases fitness of a DNA-based microbe. Proc. Natl. Acad. Sci. U. S. A **93**:906-907.
154. **Lovett M, Kuo KK, Holmes K, and Falkow S.** 1980. Plasmids of the genus *Chlamydia*, p. 1250-1252. *In* Nelson J and Grassi C (ed.), Current Chemotherapy and Infectious Diseases, vol. 2. American Society for Microbiology, Washington, DC.
155. **Clarke, I. N.** 2011. Evolution of *Chlamydia trachomatis*. Ann. N. Y. Acad. Sci. **1230**:E11-E18.

156. **Sisko, J. L., K. Spaeth, Y. Kumar, and R. H. Valdivia**. 2006. Multifunctional analysis of *Chlamydia*-specific genes in a yeast expression system. Mol. Microbiol. **60**:51-66.
157. **Sharma, J., Y. Zhong, F. Dong, J. M. Piper, G. Wang, and G. Zhong**. 2006. Profiling of human antibody responses to *Chlamydia trachomatis* urogenital tract infection using microplates arrayed with 156 chlamydial fusion proteins. Infect. Immun. **74**:1490-1499.
158. **Picard, M. D., K. P. Cohane, T. M. Gierahn, D. E. Higgins, and J. B. Flechtner**. 2012. High-throughput proteomic screening identifies *Chlamydia trachomatis* antigens that are capable of eliciting T cell and antibody responses that provide protection against vaginal challenge. Vaccine **30**:4387-4393.
159. **Starnbach, M. N., W. P. Loomis, P. Ovendale, D. Regan, B. Hess, M. R. Alderson, and S. P. Fling**. 2003. An inclusion membrane protein from *Chlamydia trachomatis* enters the MHC class I pathway and stimulates a CD8+ T cell response. J. Immunol. **171**:4742-4749.
160. **Fling, S. P., R. A. Sutherland, L. N. Steele, B. Hess, S. E. D'Orazio, J. Maisonneuve, M. F. Lampe, P. Probst, and M. N. Starnbach**. 2001. CD8+ T cells recognize an inclusion membrane-associated protein from the vacuolar pathogen *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S. A **98**:1160-1165.
161. **Follmann, F., A. W. Olsen, K. T. Jensen, P. R. Hansen, P. Andersen, and M. Theisen**. 2008. Antigenic profiling of a *Chlamydia trachomatis* gene-expression library. J. Infect. Dis. **197**:897-905.
162. **Patton, D. L., A. Teng, A. Randall, X. Liang, P. L. Felgner, and L. M. de la Maza**. 2014. Whole genome identification of C. *trachomatis* immunodominant antigens after genital tract infections and effect of antibiotic treatment of pigtailed macaques. J. Proteomics. **108**:99-109.
163. **Nunes, A., P. J. Nogueira, M. J. Borrego, and J. P. Gomes**. 2008. *Chlamydia trachomatis* diversity viewed as a tissue-specific coevolutionary arms race. Genome Biol. **9**:R153.
164. **Kari, L., W. M. Whitmire, J. H. Carlson, D. D. Crane, N. Reveneau, D. E. Nelson, D. C. Mabey, R. L. Bailey, M. J. Holland, G. McClarty, and H. D. Caldwell**. 2008. Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. J. Infect. Dis. **197**:449-456.
165. **Carlson, J. H., S. F. Porcella, G. McClarty, and H. D. Caldwell**. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. Infect. Immun. **73**:6407-6418.
166. **Seth-Smith, H. M., S. R. Harris, K. Persson, P. Marsh, A. Barron, A. Bignell, C. Bjartling, L. Clark, L. T. Cutcliffe, P. R. Lambden, N. Lennard, S. J. Lockey, M. A. Quail, O. Salim, R. J. Skilton, Y. Wang, M. J. Holland, J. Parkhill, N. R. Thomson, and I. N. Clarke**. 2009. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. BMC. Genomics **10**:239.
167. **Nunes, A., M. J. Borrego, and J. P. Gomes**. 2013. Genomic features beyond *Chlamydia trachomatis* phenotypes: what do we think we know? Infect. Genet. Evol. **16**:392-400.
168. **Brunham, R., C. Yang, I. Maclean, J. Kimani, G. Maitha, and F. Plummer**. 1994. *Chlamydia trachomatis* from individuals in a sexually transmitted disease core group exhibit frequent sequence variation in the major outer membrane protein (omp1) gene. J. Clin. Invest **94**:458-463.
169. **Gomes, J. P., W. J. Bruno, M. J. Borrego, and D. Dean**. 2004. Recombination in the genome of *Chlamydia trachomatis* involving the polymorphic membrane protein C gene relative to *ompA* and evidence for horizontal gene transfer. J. Bacteriol. **186**:4295-4306.
170. **Gomes, J. P., W. J. Bruno, A. Nunes, N. Santos, C. Florindo, M. J. Borrego, and D. Dean**. 2007. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res. **17**:50-60.
171. **Hayes, L. J., P. Yearsley, J. D. Treharne, R. A. Ballard, G. H. Fehler, and M. E. Ward**. 1994. Evidence for naturally occurring recombination in the gene encoding the major outer membrane protein of lymphogranuloma venereum isolates of *Chlamydia trachomatis*. Infect. Immun. **62**:5659-5663.
172. **Jeffrey, B. M., R. J. Suchland, K. L. Quinn, J. R. Davidson, W. E. Stamm, and D. D. Rockey**. 2010. Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect. Immun. **78**:2544-2553.
173. **Joseph, S. J., X. Didelot, J. Rothschild, H. J. de Vries, S. A. Morre, T. D. Read, and D. Dean**. 2012. Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. Mol. Biol. Evol. **29**:3933-3946.

174. **Joseph, S. J., X. Didelot, K. Gandhi, D. Dean, and T. D. Read**. 2011. Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. Biol. Direct. **6**:28.
175. **Somboonna, N., R. Wan, D. M. Ojcius, M. A. Pettengill, S. J. Joseph, A. Chang, R. Hsu, T. D. Read, and D. Dean**. 2011. Hypervirulent *Chlamydia trachomatis* clinical strain is a recombinant between lymphogranuloma venereum (L(2)) and D lineages. MBio. **2**:e00045-11.
176. **Ferreira, R., V. Borges, A. Nunes, P. J. Nogueira, M. J. Borrego, and J. P. Gomes**. 2012. Impact of loci nature on estimating recombination and mutation rates in *Chlamydia trachomatis*. G3. (Bethesda.) **2**:761-768.
177. **Bachmann, N. L., A. Polkinghorne, and P. Timms**. 2014. *Chlamydia* genomics: providing novel insights into chlamydial biology. Trends Microbiol. **22**:464-472.
178. **Byrne, G. I.** 2010. *Chlamydia trachomatis* strains and virulence: rethinking links to infection prevalence and disease severity. J. Infect. Dis. **201 Suppl 2**:S126-S133.
179. **Carlson, J. H., S. Hughes, D. Hogan, G. Cieplak, D. E. Sturdevant, G. McClarty, H. D. Caldwell, and R. J. Belland**. 2004. Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. Infect. Immun. **72**:7063-7072.
180. **Fehlner-Gardiner, C., C. Roshick, J. H. Carlson, S. Hughes, R. J. Belland, H. D. Caldwell, and G. McClarty**. 2002. Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. J. Biol. Chem. **277**:26893-26903.
181. **Wood, H., H. D. Caldwell, and G. McClarty**. 2011. Tryptophan metabolism in *Chlamydiae*, *In* Bavoil PM and Wyrick PB (ed.), *Chlamydia*: Genomics and Pathogenesis. Horizon Bioscience.
182. **Carabeo, R.** 2011. Bacterial subversion of host actin dynamics at the plasma membrane. Cell Microbiol. **13**:1460-1469.
183. **Thalmann, J., K. Janik, M. May, K. Sommer, J. Ebeling, F. Hofmann, H. Genth, and A. Klos**. 2010. Actin re-organization induced by *Chlamydia trachomatis* serovar D--evidence for a critical role of the effector protein CT166 targeting Rac. PLoS. One. **5**:e9887.
184. **Bothe, M., P. Dutow, A. Pich, H. Genth, and A. Klos**. 2015. DXD motif-dependent and -independent effects of the *Chlamydia trachomatis* cytotoxin CT166. Toxins. (Basel) **7**:621-637.
185. **McClarty, G., H. D. Caldwell, and D. E. Nelson**. 2007. Chlamydial interferon gamma immune evasion influences infection tropism. Curr. Opin. Microbiol. **10**:47-51.
186. **Lo, C. C., G. Xie, C. A. Bonner, and R. A. Jensen**. 2012. The alternative translational profile that underlies the immune-evasive state of persistence in *Chlamydia*ceae exploits differential tryptophan contents of the protein repertoire. Microbiol. Mol. Biol. Rev. **76**:405-443.
187. **Belland, R. J., M. A. Scidmore, D. D. Crane, D. M. Hogan, W. Whitmire, G. McClarty, and H. D. Caldwell**. 2001. *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proc. Natl. Acad. Sci. U. S. A **98**:13984-13989.
188. **Busch, C., F. Hofmann, R. Gerhard, and K. Aktories**. 2000. Involvement of a conserved tryptophan residue in the UDP-glucose binding of large clostridial cytotoxin glycosyltransferases. J. Biol. Chem. **275**:13228-13234.
189. **Thomson, N. R.** 2008. Vive la difference. Nat. Rev. Microbiol. **6**:502-503.
190. **Morrison, S. G., C. M. Farris, G. L. Sturdevant, W. M. Whitmire, and R. P. Morrison**. 2011. Murine *Chlamydia trachomatis* genital infection is unaltered by depletion of CD4+ T cells and diminished adaptive immunity. J. Infect. Dis. **203**:1120-1128.
191. **Hu, V. H., M. J. Holland, and M. J. Burton**. 2013. Trachoma: protective and pathogenic ocular immune responses to *Chlamydia trachomatis*. PLoS. Negl. Trop. Dis. **7**:e2020.
192. **TANG, F. F., H. L. Chang, Y. T. HUANG, and K. C. WANG**. 1957. Studies on the etiology of trachoma with special reference to isolation of the virus in chick embryo. Chin Med. J. **75**:429-447.
193. **Scidmore, M. A.** 2005. Cultivation and Laboratory Maintenance of *Chlamydia trachomatis*. Curr. Protoc. Microbiol. **Chapter 11**:Unit.
194. **Andersson, P., M. Klein, R. A. Lilliebridge, and P. M. Giffard**. 2013. Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. Clin. Microbiol. Infect. **19**:E405-E408.
195. **Putman, T. E., R. J. Suchland, J. D. Ivanovitch, and D. D. Rockey**. 2013. Culture-independent sequence analysis of *Chlamydia trachomatis* in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology **159**:2109-2117.

196. **Seth-Smith, H. M., S. R. Harris, R. J. Skilton, F. M. Radebe, D. Golparian, E. Shipitsyna, P. T. Duy, P. Scott, L. T. Cutcliffe, C. O'Neill, S. Parmar, R. Pitt, S. Baker, C. A. Ison, P. Marsh, H. Jalal, D. A. Lewis, M. Unemo, I. N. Clarke, J. Parkhill, and N. R. Thomson**. 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. Genome Res. **23**:855-866.
197. **Christiansen, M. T., A. C. Brown, S. Kundu, H. J. Tutill, R. Williams, J. R. Brown, J. Holdstock, M. J. Holland, S. Stevenson, J. Dave, C. Y. Tong, K. Einer-Jensen, D. P. Depledge, and J. Breuer**. 2014. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. BMC. Infect. Dis. **14**:591.
198. **Skinner, M. C., W. E. Stamm, and M. L. Lampe**. 2009. *Chlamydia trachomatis* laboratory strains versus recent clinical isolates: implications for routine microbicide testing. Antimicrob. Agents Chemother. **53**:1482-1489.
199. **Joubert, B. C. and A. W. Sturm**. 2011. Differences in *Chlamydia trachomatis* growth rates in human keratinocytes among lymphogranuloma venereum reference strains and clinical isolates. J. Med. Microbiol. **60**:1565-1569.
200. **Carmichael, J. R., D. Tifrea, S. Pal, and L. M. de la Maza**. 2013. Differences in infectivity and induction of infertility: a comparative study of *Chlamydia trachomatis* strains in the murine model. Microbes. Infect. **15**:219-229.
201. **Jasper, D. K., I. M. Sigar, J. H. Schripsema, C. K. Sainvil, C. L. Smith, L. Yeruva, R. G. Rank, A. K. Murthy, J. R. Widder, and K. H. Ramsey**. 2015. Genomic variant representation in a *Chlamydia* population is dynamic and adaptive with dependence on *in vitro* and *in vivo* passage. Pathog. Dis. **73**:1-12.
202. **Sturdevant, G. L., L. Kari, D. J. Gardner, N. Olivares-Zavaleta, L. B. Randall, W. M. Whitmire, J. H. Carlson, M. M. Goheen, E. M. Selleck, C. Martens, and H. D. Caldwell**. 2010. Frameshift mutations in a single novel virulence factor alter the *in vivo* pathogenicity of *Chlamydia trachomatis* for the female murine genital tract. Infect. Immun. **78**:3660-3668.
203. **Sturdevant, G. L., B. Zhou, J. H. Carlson, W. M. Whitmire, L. Song, and H. D. Caldwell**. 2014. Infectivity of urogenital *Chlamydia trachomatis* plasmid-deficient, CT135-null, and double-deficient strains in female mice. Pathog. Dis. **71**:90-92.
204. **Beare, P. A., K. M. Sandoz, A. Omsland, D. D. Rockey, and R. A. Heinzen**. 2011. Advances in genetic manipulation of obligate intracellular bacterial pathogens. Front Microbiol. **2**:97.
205. **Tam, J. E., C. H. Davis, and P. B. Wyrick**. 1994. Expression of recombinant DNA introduced into *Chlamydia trachomatis* by electroporation. Can. J. Microbiol. **40**:583-591.
206. **Wang, Y., S. Kahane, L. T. Cutcliffe, R. J. Skilton, P. R. Lambden, and I. N. Clarke**. 2011. Development of a transformation system for *Chlamydia trachomatis*: restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. PLoS. Pathog. **7**:e1002258.
207. **Song, L., J. H. Carlson, W. M. Whitmire, L. Kari, K. Virtaneva, D. E. Sturdevant, H. Watkins, B. Zhou, G. L. Sturdevant, S. F. Porcella, G. McClarty, and H. D. Caldwell**. 2013. *Chlamydia trachomatis* plasmid-encoded Pgp4 is a transcriptional regulator of virulence-associated genes. Infect. Immun. **81**:636-644.
208. **Peterson, E. M., B. A. Markoff, J. Schachter, and L. M. de la Maza**. 1990. The 7.5-kb plasmid present in *Chlamydia trachomatis* is not essential for the growth of this microorganism. Plasmid **23**:144-148.
209. **Farencena, A., M. Comanducci, M. Donati, G. Ratti, and R. Cevenini**. 1997. Characterization of a new isolate of *Chlamydia trachomatis* which lacks the common plasmid and has properties of biovar trachoma. Infect. Immun. **65**:2965-2969.
210. **Matsumoto, A., H. Izutsu, N. Miyashita, and M. Ohuchi**. 1998. Plaque formation by and plaque cloning of *Chlamydia trachomatis* biovar trachoma. J. Clin. Microbiol. **36**:3013-3019.
211. **Stothard, D. R., J. A. Williams, P. B. Van Der, and R. B. Jones**. 1998. Identification of a *Chlamydia trachomatis* serovar E urogenital isolate which lacks the cryptic plasmid. Infect. Immun. **66**:6010-6013.
212. **Agaisse, H. and I. Derre**. 2013. A C. *trachomatis* cloning vector and the generation of C. *trachomatis* strains expressing fluorescent proteins under the control of a C. *trachomatis* promoter. PLoS. One. **8**:e57090.
213. **Kari, L., M. M. Goheen, L. B. Randall, L. D. Taylor, J. H. Carlson, W. M. Whitmire, D. Virok, K. Rajaram, V. Endresz, G. McClarty, D. E. Nelson, and H. D. Caldwell**. 2011. Generation of targeted *Chlamydia trachomatis* null mutants. Proc. Natl. Acad. Sci. U. S. A **108**:7189-7193.

214. **Johnson, C. M. and D. J. Fisher**. 2013. Site-specific, insertional inactivation of incA in *Chlamydia trachomatis* using a group II intron. PLoS. One. **8**:e83989.
215. **Gerard, H. C., M. K. Mishra, G. Mao, S. Wang, M. Hali, J. A. Whittum-Hudson, R. M. Kannan, and A. P. Hudson**. 2013. Dendrimer-enabled DNA delivery and transformation of *Chlamydia pneumoniae*. Nanomedicine. **9**:996-1008.
216. **Mishra, M. K., H. C. Gerard, J. A. Whittum-Hudson, A. P. Hudson, and R. M. Kannan**. 2012. Dendrimer-enabled modulation of gene expression in *Chlamydia trachomatis*. Mol. Pharm. **9**:413-421.
217. **Nguyen, B. D. and R. H. Valdivia**. 2012. Virulence determinants in the obligate intracellular pathogen *Chlamydia trachomatis* revealed by forward genetic approaches. Proc. Natl. Acad. Sci. U. S. A **109**:1263-1268.
218. **Wang, Y., S. Kahane, L. T. Cutcliffe, R. J. Skilton, P. R. Lambden, K. Persson, C. Bjartling, and I. N. Clarke**. 2013. Genetic transformation of a clinical (genital tract), plasmid-free isolate of *Chlamydia trachomatis*: engineering the plasmid as a cloning vector. PLoS. One. **8**:e59195.
219. **Wickstrum, J., L. R. Sammons, K. N. Restivo, and P. S. Hefty**. 2013. Conditional gene expression in *Chlamydia trachomatis* using the tet system. PLoS. One. **8**:e76743.
220. **Kokes, M., J. D. Dunn, J. A. Granek, B. D. Nguyen, J. R. Barker, R. H. Valdivia, and R. J. Bastidas**. 2015. Integrating Chemical Mutagenesis and Whole-Genome Sequencing as a Platform for Forward and Reverse Genetic Analysis of *Chlamydia*. Cell Host. Microbe .
221. **Du, T. A.** 2015. Techniques &applications: Expanding the toolbox to study *Chlamydia*. Nat. Rev. Microbiol.
222. **de, B. M. and G. Greub**. 2013. Functional genomics of intracellular bacteria. Brief. Funct. Genomics **12**:341-353.
223. **Subtil, A. and A. utry-Varsat**. 2004. *Chlamydia*: five years A.G. (after genome). Curr. Opin. Microbiol. **7**:85-92.
224. **Fraser, C., W. P. Hanage, and B. G. Spratt**. 2005. Neutral microepidemic evolution of bacterial pathogens. Proc. Natl. Acad. Sci. U. S. A **102**:1968-1973.
225. **Bush, R. M.** 2001. Predicting adaptive evolution. Nat. Rev. Genet. **2**:387-392.
226. **Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein**. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. **8**:845-856.
227. **Chen, L., A. Perlina, and C. J. Lee**. 2004. Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. J. Virol. **78**:3722-3732.
228. **Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T. Bhattacharya, and B. Korber**. 2002. Diversity considerations in HIV-1 vaccine selection. Science **296**:2354-2360.
229. **Bush, R. M., C. A. Bender, K. Subbarao, N. J. Cox, and W. M. Fitch**. 1999. Predicting the evolution of human influenza A. Science **286**:1921-1925.
230. **Anisimova, M., J. Bielawski, K. Dunn, and Z. Yang**. 2007. Phylogenomic analysis of natural selection pressure in Streptococcus genomes. BMC. Evol. Biol. **7**:154.
231. **Borges, V., R. Ferreira, A. Nunes, P. Nogueira, M. J. Borrego, and J. P. Gomes**. 2010. Normalization strategies for real-time expression data in *Chlamydia trachomatis*. J. Microbiol. Methods **82**:256-264.
232. **Darling, A. E., B. Mau, and N. T. Perna**. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS. One. **5**:e11147.
233. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731-2739.
234. **Saitou, N. and M. Nei**. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406-425.
235. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111-120.
236. **Yang, Z. and R. Nielsen**. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**:32-43.
237. **Nei M and Kumar S**. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York, NY.
238. **Egea, R., S. Casillas, and A. Barbadilla**. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res. **36**:W157-W162.

239. **McDonald, J. H. and M. Kreitman**. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature **351**:652-654.
240. **Nielsen, R.** 2001. Statistical tests of selective neutrality in the age of genomics. Heredity (Edinb.) **86**:641-647.
241. **Yang, Z. and R. Nielsen**. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19**:908-917.
242. **Zhang, J., R. Nielsen, and Z. Yang**. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. **22**:2472-2479.
243. **Yang, Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586-1591.
244. **Yang, Z. and R. M. dos**. 2011. Statistical properties of the branch-site test of positive selection. Mol. Biol. Evol. **28**:1217-1228.
245. **Yang, Z.** 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568-573.
246. **Yang, Z., W. S. Wong, and R. Nielsen**. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22**:1107-1118.
247. **Carver, T., N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill**. 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. **25**:119-120.
248. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambridge University Press, London, England.
249. **Jordan, I. K., I. B. Rogozin, Y. I. Wolf, and E. V. Koonin**. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. **12**:962-968.
250. **Tan, C., R. C. Hsia, H. Shou, C. L. Haggerty, R. B. Ness, C. A. Gaydos, D. Dean, A. M. Scurlock, D. P. Wilson, and P. M. Bavoil**. 2009. *Chlamydia trachomatis*-infected patients display variable antibody profiles against the nine-member polymorphic membrane protein family. Infect. Immun. **77**:3218-3226.
251. **Gomes, J. P., R. C. Hsia, S. Mead, M. J. Borrego, and D. Dean**. 2005. Immunoreactivity and differential developmental expression of known and putative *Chlamydia trachomatis* membrane proteins for biologically variant serovars representing distinct disease groups. Microbes. Infect. **7**:410-420.
252. **Coler, R. N., A. Bhatia, J. F. Maisonneuve, P. Probst, B. Barth, P. Ovendale, H. Fang, M. Alderson, Y. Lobet, J. Cohen, P. Mettens, and S. G. Reed**. 2009. Identification and characterization of novel recombinant vaccine antigens for immunization against genital *Chlamydia trachomatis*. FEMS Immunol. Med. Microbiol. **55**:258-270.
253. **Scidmore-Carlson, M. A., E. I. Shaw, C. A. Dooley, E. R. Fischer, and T. Hackstadt**. 1999. Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. Mol. Microbiol. **33**:753-765.
254. **Rockey, D. D., W. Viratyosin, J. P. Bannantine, R. J. Suchland, and W. E. Stamm**. 2002. Diversity within inc genes of clinical *Chlamydia trachomatis* variant isolates that occupy non-fusogenic inclusions. Microbiology **148**:2497-2505.
255. **Delevoye, C., M. Nilges, P. Dehoux, F. Paumet, S. Perrinet, A. utry-Varsat, and A. Subtil**. 2008. SNARE protein mimicry by an intracellular bacterium. PLoS. Pathog. **4**:e1000022.
256. **Jewett, T. J., C. A. Dooley, D. J. Mead, and T. Hackstadt**. 2008. *Chlamydia trachomatis* tarp is phosphorylated by src family tyrosine kinases. Biochem. Biophys. Res. Commun. **371**:339-344.
257. **Jorgensen, I. and R. H. Valdivia**. 2008. Pmp-like proteins Pls1 and Pls2 are secreted into the lumen of the *Chlamydia trachomatis* inclusion. Infect. Immun. **76**:3940-3950.
258. **Mital, J., N. J. Miller, E. R. Fischer, and T. Hackstadt**. 2010. Specific chlamydial inclusion membrane proteins associate with active Src family kinases in microdomains that interact with the host microtubule network. Cell Microbiol. **12**:1235-1249.
259. **Bannantine, J. P., R. S. Griffiths, W. Viratyosin, W. J. Brown, and D. D. Rockey**. 2000. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. Cell Microbiol. **2**:35-47.
260. **Alzhanov, D. T., S. K. Weeks, J. R. Burnett, and D. D. Rockey**. 2009. Cytokinesis is blocked in mammalian cells transfected with *Chlamydia trachomatis* gene CT223. BMC. Microbiol. **9**:2.
261. **Bannantine, J. P., D. D. Rockey, and T. Hackstadt**. 1998. Tandem genes of *Chlamydia psittaci* that encode proteins localized to the inclusion membrane. Mol. Microbiol. **28**:1017-1026.

262. **Jia, T. J., D. W. Liu, J. H. Luo, and G. M. Zhong**. 2007. [Localization of the hypothetical protein CT249 in the *Chlamydia trachomatis* inclusion membrane]. Wei Sheng Wu Xue. Bao. **47**:645-648.
263. **Wang, J., L. Chen, F. Chen, X. Zhang, Y. Zhang, J. Baseman, S. Perdue, I. T. Yeh, R. Shain, M. Holland, R. Bailey, D. Mabey, P. Yu, and G. Zhong**. 2009. A chlamydial type III-secreted effector protein (Tarp) is predominantly recognized by antibodies from humans infected with *Chlamydia trachomatis* and induces protective immunity against upper genital tract pathologies in mice. Vaccine **27**:2967-2980.
264. **Finco, O., E. Frigimelica, F. Buricchi, R. Petracca, G. Galli, E. Faenzi, E. Meoni, A. Bonci, M. Agnusdei, F. Nardelli, E. Bartolini, M. Scarselli, E. Caproni, D. Laera, L. Zedda, D. Skibinski, S. Giovinazzi, R. Bastone, E. Ianni, R. Cevenini, G. Grandi, and R. Grifantini**. 2011. Approach to discover T- and B-cell antigens of intracellular pathogens applied to the design of *Chlamydia trachomatis* vaccines. Proc. Natl. Acad. Sci. U. S. A **108**:9969-9974.
265. **Gong, S., L. Lei, X. Chang, R. Belland, and G. Zhong**. 2011. *Chlamydia trachomatis* secretion of hypothetical protein CT622 into host cell cytoplasm via a secretion pathway that can be inhibited by the type III secretion system inhibitor compound 1. Microbiology **157**:1134-1144.
266. **Hower, S., K. Wolf, and K. A. Fields**. 2009. Evidence that CT694 is a novel *Chlamydia trachomatis* T3S substrate capable of functioning during invasion or early cycle development. Mol. Microbiol. **72**:1423-1437.
267. **Bullock, H. D., S. Hower, and K. A. Fields**. 2012. Domain analyses reveal that *Chlamydia trachomatis* CT694 protein belongs to the membrane-localized family of type III effector proteins. J. Biol. Chem. **287**:28078-28086.
268. **Misaghi, S., Z. R. Balsara, A. Catic, E. Spooner, H. L. Ploegh, and M. N. Starnbach**. 2006. *Chlamydia trachomatis*-derived deubiquitinating enzymes in mammalian cells during infection. Mol. Microbiol. **61**:142-150.
269. **Le, N. G., A. Krieg, B. Faustin, M. Loeffler, A. Godzik, S. Krajewski, and J. C. Reed**. 2008. ChlaDub1 of *Chlamydia trachomatis* suppresses NF-kappaB activation and inhibits IkappaBalpha ubiquitination and degradation. Cell Microbiol. **10**:1879-1892.
270. **Stephens, R. S.** 2002. *Chlamydiae* and evolution:a billion years and counting., p. 3-12. *In* C. H. K. K. R. R. S. S. S. S. T. P. &. W. (Schachter C (ed.), Chlamydial Infections: Proceedings of the Tenth International Symposium on Human chlamydial Infections. International *Chlamydia*l Symposium, San Francisco, CA.
271. **Anisimova, M. and Z. Yang**. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol. Biol. Evol. **24**:1219-1228.
272. **Vos, M.** 2009. Why do bacteria engage in homologous recombination? Trends Microbiol. **17**:226-232.
273. **Gomes, J. P., M. J. Borrego, B. Atik, I. Santo, J. Azevedo, S. A. Brito de, P. Nogueira, and D. Dean**. 2006. Correlating *Chlamydia trachomatis* infectious load with urogenital ecological success and disease pathogenesis. Microbes. Infect. **8**:16-26.
274. **Chao, L.** 1990. Fitness of RNA virus decreased by Muller's ratchet. Nature **348**:454-455.
275. **Jewett, T. J., E. R. Fischer, D. J. Mead, and T. Hackstadt**. 2006. Chlamydial TARP is a bacterial nucleator of actin. Proc. Natl. Acad. Sci. U. S. A **103**:15599-15604.
276. **Mehlitz, A., S. Banhart, A. P. Maurer, A. Kaushansky, A. G. Gordus, J. Zielecki, G. Macbeath, and T. F. Meyer**. 2010. Tarp regulates early *Chlamydia*-induced host cell survival through interactions with the human adaptor protein SHC1. J. Cell Biol. **190**:143-157.
277. **Nunes, A., J. P. Gomes, S. Mead, C. Florindo, H. Correia, M. J. Borrego, and D. Dean**. 2007. Comparative expression profiling of the *Chlamydia trachomatis* pmp gene family for clinical and reference strains. PLoS. One. **2**:e878.
278. **Tan, C., R. C. Hsia, H. Shou, J. A. Carrasco, R. G. Rank, and P. M. Bavoil**. 2010. Variable expression of surface-exposed polymorphic membrane proteins in *in vitro*-grown *Chlamydia trachomatis*. Cell Microbiol. **12**:174-187.
279. **Jordan, I. K., F. A. Kondrashov, I. B. Rogozin, R. L. Tatusov, Y. I. Wolf, and E. V. Koonin**. 2001. Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. Genome Biol. **2**:RESEARCH0053.
280. **Cambray, G. and D. Mazel**. 2008. Synonymous genes explore different evolutionary landscapes. PLoS. Genet. **4**:e1000256.

281. **Chamary, J. V. and L. D. Hurst**. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol. **6**:R75.
282. **Sharp, P. M., T. M. Tuohy, and K. R. Mosurski**. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. **14**:5125-5143.
283. **Kanaya, S., Y. Yamada, Y. Kudo, and T. Ikemura**. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene **238**:143-155.
284. **Unemo, M., H. M. Seth-Smith, L. T. Cutcliffe, R. J. Skilton, D. Barlow, D. Goulding, K. Persson, S. R. Harris, A. Kelly, C. Bjartling, H. Fredlund, P. Olcen, N. R. Thomson, and I. N. Clarke**. 2010. The Swedish new variant of *Chlamydia trachomatis*: genome sequence, morphology, cell tropism and phenotypic characterization. Microbiology **156**:1394-1404.
285. **Shaw, A. C., G. Christiansen, P. Roepstorff, and S. Birkelund**. 2000. Genetic differences in the *Chlamydia trachomatis* tryptophan synthase alpha-subunit can explain variations in serovar pathogenesis. Microbes. Infect. **2**:581-592.
286. **Nelson, D. E., D. D. Crane, L. D. Taylor, D. W. Dorward, M. M. Goheen, and H. D. Caldwell**. 2006. Inhibition of *Chlamydiae* by primary alcohols correlates with the strain-specific complement of plasticity zone phospholipase D genes. Infect. Immun. **74**:73-80.
287. **Lutter, E. I., C. Bonner, M. J. Holland, R. J. Suchland, W. E. Stamm, T. J. Jewett, G. McClarty, and T. Hackstadt**. 2010. Phylogenetic analysis of *Chlamydia trachomatis* Tarp and correlation with clinical phenotype. Infect. Immun. **78**:3678-3688.
288. **Ho, T. D. and M. N. Starnbach**. 2005. The *Salmonella enterica* serovar typhimurium-encoded type III secretion systems can translocate *Chlamydia trachomatis* proteins into the cytosol of host cells. Infect. Immun. **73**:905-911.
289. **Subtil, A., C. Parsot, and A. utry-Varsat**. 2001. Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. Mol. Microbiol. **39**:792-800.
290. **Chellas-Gery, B., C. N. Linton, and K. A. Fields**. 2007. Human GCIP interacts with CT847, a novel *Chlamydia trachomatis* type III secretion substrate, and is degraded in a tissue-culture infection model. Cell Microbiol. **9**:2417-2430.
291. **Fields, K. A. and T. Hackstadt**. 2000. Evidence for the secretion of *Chlamydia trachomatis* CopN by a type III secretion mechanism. Mol. Microbiol. **38**:1048-1060.
292. **Lutter, E. I., C. Martens, and T. Hackstadt**. 2012. Evolution and conservation of predicted inclusion membrane proteins in *Chlamydiae*. Comp Funct. Genomics **2012**:362104.
293. **Derre, I., R. Swiss, and H. Agaisse**. 2011. The lipid transfer protein CERT interacts with the *Chlamydia* inclusion protein IncD and participates to ER-*Chlamydia* inclusion membrane contact sites. PLoS. Pathog. **7**:e1002092.
294. **Rzomp, K. A., A. R. Moorhead, and M. A. Scidmore**. 2006. The GTPase Rab4 interacts with *Chlamydia trachomatis* inclusion membrane protein CT229. Infect. Immun. **74**:5362-5373.
295. **Iriarte, M. and G. R. Cornelis**. 1998. YopT, a new Yersinia Yop effector protein, affects the cytoskeleton of host cells. Mol. Microbiol. **29**:915-929.
296. **Sorg, I., S. Wagner, M. Amstutz, S. A. Muller, P. Broz, Y. Lussi, A. Engel, and G. R. Cornelis**. 2007. YscU recognizes translocators as export substrates of the Yersinia injectisome. EMBO J. **26**:3015-3024.
297. **Charpentier, X. and E. Oswald**. 2004. Identification of the secretion and translocation domain of the enteropathogenic and enterohemorrhagic *Escherichia coli* effector Cif, using TEM-1 beta-lactamase as a new fluorescence-based reporter. J. Bacteriol. **186**:5486-5495.
298. **Marenne, M. N., L. Journet, L. J. Mota, and G. R. Cornelis**. 2003. Genetic analysis of the formation of the Ysc-Yop translocation pore in macrophages by Yersinia enterocolitica: role of LcrV, YscF and YopN. Microb. Pathog. **35**:243-258.
299. **Letzelter, M., I. Sorg, L. J. Mota, S. Meyer, J. Stalder, M. Feldman, M. Kuhn, I. Callebaut, and G. R. Cornelis**. 2006. The discovery of SycO highlights a new function for type III secretion effector chaperones. EMBO J. **25**:3223-3233.
300. **Li, Z., C. Chen, D. Chen, Y. Wu, Y. Zhong, and G. Zhong**. 2008. Characterization of fifty putative inclusion membrane proteins encoded in the *Chlamydia trachomatis* genome. Infect. Immun. **76**:2746-2757.
301. **Borges, V., A. Nunes, R. Ferreira, M. J. Borrego, and J. P. Gomes**. 2012. Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation. J. Bacteriol. **194**:6143-6153.

302. **Brunelle, B. W., T. L. Nicholson, and R. S. Stephens**. 2004. Microarray-based genomic surveying of gene polymorphisms in *Chlamydia trachomatis*. Genome Biol. **5**:R42.
303. **Papenfort, K., D. Podkaminski, J. C. Hinton, and J. Vogel**. 2012. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. Proc. Natl. Acad. Sci. U. S. A **109**:E757-E764.
304. **Yong, E. C., E. Y. Chi, and C. C. Kuo**. 1987. Differential antimicrobial activity of human mononuclear phagocytes against the human biovars of *Chlamydia trachomatis*. J. Immunol. **139**:1297-1302.
305. **Flannagan, R. S., V. Jaumouille, and S. Grinstein**. 2012. The cell biology of phagocytosis. Annu. Rev. Pathol. **7**:61-98.
306. **Lane, B. J., C. Mutchler, K. S. Al, S. S. Grieshaber, and R. A. Carabeo**. 2008. Chlamydial entry involves TARP binding of guanine nucleotide exchange factors. PLoS. Pathog. **4**:e1000014.
307. **Lutter, E. I., A. C. Barger, V. Nair, and T. Hackstadt**. 2013. *Chlamydia trachomatis* inclusion membrane protein CT228 recruits elements of the myosin phosphatase pathway to regulate release mechanisms. Cell Rep. **3**:1921-1931.
308. **Scidmore, M. A. and T. Hackstadt**. 2001. Mammalian 14-3-3beta associates with the *Chlamydia trachomatis* inclusion membrane via its interaction with IncG. Mol. Microbiol. **39**:1638-1650.
309. **Pennini, M. E., S. Perrinet, A. utry-Varsat, and A. Subtil**. 2010. Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis*. PLoS. Pathog. **6**:e1000995.
310. **Muschiol, S., G. Boncompain, F. Vromman, P. Dehoux, S. Normark, B. Henriques-Normark, and A. Subtil**. 2011. Identification of a family of effectors secreted by the type III secretion system that are conserved in pathogenic *Chlamydiae*. Infect. Immun. **79**:571-580.
311. **Furtado, A. R., M. Essid, S. Perrinet, M. E. Balana, N. Yoder, P. Dehoux, and A. Subtil**. 2013. The chlamydial OTU domain-containing protein ChlaOTU is an early type III secretion effector targeting ubiquitin and NDP52. Cell Microbiol. **15**:2064-2079.
312. **Pais, S. V., C. Milho, F. Almeida, and L. J. Mota**. 2013. Identification of novel type III secretion chaperone-substrate complexes of *Chlamydia trachomatis*. PLoS. One. **8**:e56292.
313. **Hovis, K. M., S. Mojica, J. E. McDermott, L. Pedersen, C. Simhi, R. G. Rank, G. S. Myers, J. Ravel, R. C. Hsia, and P. M. Bavoil**. 2013. Genus-optimized strategy for the identification of chlamydial type III secretion substrates. Pathog. Dis. **69**:213-222.
314. **Arnold, R., S. Brandmaier, F. Kleine, P. Tischler, E. Heinz, S. Behrens, A. Niinikoski, H. W. Mewes, M. Horn, and T. Rattei**. 2009. Sequence-based prediction of type III secreted proteins. PLoS. Pathog. **5**:e1000376.
315. **Samudrala, R., F. Heffron, and J. E. McDermott**. 2009. Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. PLoS. Pathog. **5**:e1000375.
316. **Lower, M. and G. Schneider**. 2009. Prediction of type III secretion signals in genomes of gram-negative bacteria. PLoS. One. **4**:e5917.
317. **Hobolt-Pedersen, A. S., G. Christiansen, E. Timmerman, K. Gevaert, and S. Birkelund**. 2009. Identification of *Chlamydia trachomatis* CT621, a protein delivered through the type III secretion system to the host cell cytoplasm and nucleus. FEMS Immunol. Med. Microbiol. **57**:46-58.
318. **Kumar, Y., J. Cocchiaro, and R. H. Valdivia**. 2006. The obligate intracellular pathogen *Chlamydia trachomatis* targets host lipid droplets. Curr. Biol. **16**:1646-1651.
319. **Lei, L., M. Qi, N. Budrys, R. Schenken, and G. Zhong**. 2011. Localization of *Chlamydia trachomatis* hypothetical protein CT311 in host cell cytoplasm. Microb. Pathog. **51**:101-109.
320. **Qi, M., L. Lei, S. Gong, Q. Liu, M. P. DeLisa, and G. Zhong**. 2011. *Chlamydia trachomatis* secretion of an immunodominant hypothetical protein (CT795) into host cell cytoplasm. J. Bacteriol. **193**:2498-2509.
321. **Lu, C., L. Lei, B. Peng, L. Tang, H. Ding, S. Gong, Z. Li, Y. Wu, and G. Zhong**. 2013. *Chlamydia trachomatis* GlgA is secreted into host cell cytoplasm. PLoS. One. **8**:e68764.
322. **Li, Z., D. Chen, Y. Zhong, S. Wang, and G. Zhong**. 2008. The chlamydial plasmid-encoded protein pgp3 is secreted into the cytosol of *Chlamydia*-infected cells. Infect. Immun. **76**:3415-3428.
323. **Lei, L., X. Dong, Z. Li, and G. Zhong**. 2013. Identification of a novel nuclear localization signal sequence in *Chlamydia trachomatis*-secreted hypothetical protein CT311. PLoS. One. **8**:e64529.

324. **Almeida, F., V. Borges, R. Ferreira, M. J. Borrego, J. P. Gomes, and L. J. Mota**. 2012. Polymorphisms in inc proteins and differential expression of *inc* genes among *Chlamydia trachomatis* strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates. J. Bacteriol. **194**:6574-6585.
325. **Denecker, G., S. Totemeyer, L. J. Mota, P. Troisfontaines, I. Lambermont, C. Youta, I. Stainier, M. Ackermann, and G. R. Cornelis**. 2002. Effect of low- and high-virulence *Yersinia enterocolitica* strains on the inflammatory response of human umbilical vein endothelial cells. Infect. Immun. **70**:3510-3520.
326. **Grosdent, N., I. Maridonneau-Parini, M. P. Sory, and G. R. Cornelis**. 2002. Role of Yops and adhesins in resistance of *Yersinia enterocolitica* to phagocytosis. Infect. Immun. **70**:4165-4176.
327. **Bao, X., B. E. Nickels, and H. Fan**. 2012. *Chlamydia trachomatis* protein GrgA activates transcription by contacting the nonconserved region of sigma66. Proc. Natl. Acad. Sci. U. S. A **109**:16870-16875.
328. **Wang, Y., M. Sun, H. Bao, and A. P. White**. 2013. T3_MM: a Markov model effectively classifies bacterial type III secretion signals. PLoS. One. **8**:e58173.
329. **Sory, M. P., A. Boland, I. Lambermont, and G. R. Cornelis**. 1995. Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach. Proc. Natl. Acad. Sci. U. S. A **92**:11998-12002.
330. **Lloyd, S. A., M. Norman, R. Rosqvist, and H. Wolf-Watz**. 2001. Yersinia YopE is targeted for type III secretion by N-terminal, not mRNA, signals. Mol. Microbiol. **39**:520-531.
331. **Feldman, M. F., S. Muller, E. Wuest, and G. R. Cornelis**. 2002. SycE allows secretion of YopE-DHFR hybrids by the *Yersinia enterocolitica* type III Ysc system. Mol. Microbiol. **46**:1183-1197.
332. **Lee, V. T. and O. Schneewind**. 2002. Yop fusions to tightly folded protein domains and their effects on *Yersinia enterocolitica* type III secretion. J. Bacteriol. **184**:3740-3745.
333. **Akeda, Y. and J. E. Galan**. 2005. Chaperone release and unfolding of substrates in type III secretion. Nature **437**:911-915.
334. **Sorg, J. A., N. C. Miller, M. M. Marketon, and O. Schneewind**. 2005. Rejection of impassable substrates by *Yersinia* type III secretion machines. J. Bacteriol. **187**:7090-7102.
335. **Stebbins, C. E. and J. E. Galan**. 2001. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. Nature **414**:77-81.
336. **Rockey, D. D.** 2011. Unraveling the basic biology and clinical significance of the chlamydial plasmid. J. Exp. Med. **208**:2159-2162.
337. **Kari, L., W. M. Whitmire, N. Olivares-Zavaleta, M. M. Goheen, L. D. Taylor, J. H. Carlson, G. L. Sturdevant, C. Lu, L. E. Bakios, L. B. Randall, M. J. Parnell, G. Zhong, and H. D. Caldwell**. 2011. A live-attenuated chlamydial vaccine protects against trachoma in nonhuman primates. J. Exp. Med. **208**:2217-2223.
338. **Olivares-Zavaleta, N., W. Whitmire, D. Gardner, and H. D. Caldwell**. 2010. Immunization with the attenuated plasmidless *Chlamydia trachomatis* L2(25667R) strain provides partial protection in a murine model of female genitourinary tract infection. Vaccine **28**:1454-1462.
339. **Spaeth, K. E., Y. S. Chen, and R. H. Valdivia**. 2009. The *Chlamydia* type III secretion system C-ring engages a chaperone-effector protein complex. PLoS. Pathog. **5**:e1000579.
340. **Ponting, C. P.** 1999. Chlamydial homologues of the MACPF (MAC/perforin) domain. Curr. Biol. **9**:R911-R913.
341. **Taylor, L. D., D. E. Nelson, D. W. Dorward, W. M. Whitmire, and H. D. Caldwell**. 2010. Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. Infect. Immun. **78**:2691-2699.
342. **Pettersson, J., R. Nordfelth, E. Dubinina, T. Bergman, M. Gustafsson, K. E. Magnusson, and H. Wolf-Watz**. 1996. Modulation of virulence factor expression by pathogen target cell contact. Science **273**:1231-1233.
343. **Parsot, C., E. Ageron, C. Penno, M. Mavris, K. Jamoussi, H. d'Hauteville, P. Sansonetti, and B. Demers**. 2005. A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*. Mol. Microbiol. **56**:1627-1635.
344. **Botteaux, A., M. P. Sory, L. Biskri, C. Parsot, and A. Allaoui**. 2009. MxiC is secreted by and controls the substrate specificity of the *Shigella flexneri* type III secretion apparatus. Mol. Microbiol. **71**:449-460.

345. **Feldman, M. F. and G. R. Cornelis**. 2003. The multitalented type III chaperones: all you can do with 15 kDa. FEMS Microbiol. Lett. **219**:151-158.
346. **Parsot, C., C. Hamiaux, and A. L. Page**. 2003. The various and varying roles of specific chaperones in type III secretion systems. Curr. Opin. Microbiol. **6**:7-14.
347. **World Health Organization**. 2012. Global WHO alliance for the elimination of blinding trachoma by 2020. Wkly. Epidemiol. Rec. WHO, Switzerland, Geneva. **87**:161-168.
348. **Stevens, M. P., S. N. Tabrizi, R. Muller, V. Krause, and S. M. Garland**. 2004. Characterization of *Chlamydia trachomatis* omp1 genotypes detected in eye swab samples from remote Australian communities. J. Clin. Microbiol. **42**:2501-2507.
349. **Grayston, J. T., S. P. Wang, R. L. WOOLRIDGE, Y. F. YANG, and P. B. JOHNSTON**. 1960. Trachoma-studies of etiology, laboratory diagnosis, and prevention. J. Am. Med. Assoc. **172**:1577-1586.
350. **Borges, V., R. Ferreira, A. Nunes, M. Sousa-Uva, M. Abreu, M. J. Borrego, and J. P. Gomes**. 2013. Effect of long-term laboratory propagation on *Chlamydia trachomatis* genome dynamics. Infect. Genet. Evol. **17**:23-32.
351. **Langmead, B. and S. L. Salzberg**. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357-359.
352. **Li, H. and R. Durbin**. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. **26**:589-595.
353. **Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. **25**:2078-2079.
354. **Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov**. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. **14**:178-192.
355. **Ferreira, R., V. Borges, A. Nunes, M. J. Borrego, and J. P. Gomes**. 2013. Assessment of the load and transcriptional dynamics of *Chlamydia trachomatis* plasmid according to strains' tissue tropism. Microbiol. Res. **168**:333-339.
356. **Giles, T. N., D. J. Fisher, and D. E. Graham**. 2009. Independent inactivation of arginine decarboxylase genes by nonsense and missense mutations led to pseudogene formation in *Chlamydia trachomatis* serovar L2 and D strains. BMC. Evol. Biol. **9**:166.
357. **Borges, V., A. Nunes, R. Ferreira, M. J. Borrego, and J. P. Gomes**. 2012. Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation. J. Bacteriol. **194**:6143-6153.
358. **Bauwens, J. E., H. Orlander, M. P. Gomez, M. Lampe, S. Morse, W. E. Stamm, R. Cone, R. Ashley, P. Swenson, and K. K. Holmes**. 2002. Epidemic Lymphogranuloma venereum during epidemics of crack cocaine use and HIV infection in the Bahamas. Sex Transm. Dis. **29**:253-259.
359. **Behets, F. M., J. Andriamiadana, D. Randrianasolo, R. Randriamanga, D. Rasamilalao, C. Y. Chen, J. B. Weiss, S. A. Morse, G. Dallabetta, and M. S. Cohen**. 1999. Chancroid, primary syphilis, genital herpes, and lymphogranuloma venereum in Antananarivo, Madagascar. J. Infect. Dis. **180**:1382-1385.
360. **O'Farrell, N., A. A. Hoosen, K. D. Coetzee, and E. J. van den**. 1991. Genital ulcer disease in men in Durban, South Africa. Genitourin. Med. **67**:327-330.
361. **Viravan, C., D. A. Dance, C. Ariyarit, S. Looareesuwan, Y. Wattanagoon, T. M. Davis, V. Wuthiekanun, S. Tantivanich, B. J. Angus, and N. J. White**. 1996. A prospective clinical and bacteriologic study of inguinal buboes in Thai men. Clin. Infect. Dis. **22**:233-239.
362. **de Vrieze, N. H. and H. J. de Vries**. 2014. Lymphogranuloma venereum among men who have sex with men. An epidemiological and clinical review. Expert. Rev. Anti. Infect. Ther. **12**:697-704.
363. **van de Laar, M. J.** 2006. The emergence of LGV in western Europe: what do we know, what can we do? Euro. Surveill **11**:146-148.
364. **Librado, P. and J. Rozas**. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. **25**:1451-1452.
365. **Ferreira, R., M. Antelo, A. Nunes, V. Borges, V. Damiao, M. J. Borrego, and J. P. Gomes**. 2014. In silico scrutiny of genes revealing phylogenetic congruence with clinical prevalence or tropism properties of *Chlamydia trachomatis* strains. G3. (Bethesda.) **5**:9-19.
366. **Romero, H., A. Zavala, and H. Musto**. 2000. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res. **28**:2084-2090.

367. **Kuo, C. H., N. A. Moran, and H. Ochman**. 2009. The consequences of genetic drift for bacterial genome complexity. Genome Res. **19**:1450-1454.
368. **Chen, Y. S., R. J. Bastidas, H. A. Saka, V. K. Carpenter, K. L. Richards, G. V. Plano, and R. H. Valdivia**. 2014. The *Chlamydia trachomatis* type III secretion chaperone Slc1 engages multiple early effectors, including TepP, a tyrosine-phosphorylated protein required for the recruitment of CrkI-II to nascent inclusions and innate immune signaling. PLoS. Pathog. **10**:e1003954.
369. **da, Cunha M., C. Milho, F. Almeida, S. V. Pais, V. Borges, R. Mauricio, M. J. Borrego, J. P. Gomes, and L. J. Mota**. 2014. Identification of type III secretion substrates of *Chlamydia trachomatis* using *Yersinia enterocolitica* as a heterologous system. BMC. Microbiol. **14**:40.
370. **Agaisse, H. and I. Derre**. 2014. Expression of the effector protein IncD in *Chlamydia trachomatis* mediates recruitment of the lipid transfer protein CERT and the endoplasmic reticulum-resident protein VAPB to the inclusion membrane. Infect. Immun. **82**:2037-2047.
371. **Verbeke, P., L. Welter-Stahl, S. Ying, J. Hansen, G. Hacker, T. Darville, and D. M. Ojcius**. 2006. Recruitment of BAD by the *Chlamydia trachomatis* vacuole correlates with host-cell survival. PLoS. Pathog. **2**:e45.
372. **Gupta, R., P. Srivastava, H. Vardhan, S. Salhan, and A. Mittal**. 2009. Host immune responses to chlamydial inclusion membrane proteins B and C in *Chlamydia trachomatis* infected women with or without fertility disorders. Reprod. Biol. Endocrinol. **7**:38.
373. **Gupta, R., H. Vardhan, P. Srivastava, S. Salhan, and A. Mittal**. 2009. Modulation of cytokines and transcription factors (T-Bet and GATA3) in CD4 enriched cervical cells of *Chlamydia trachomatis* infected fertile and infertile women upon stimulation with chlamydial inclusion membrane proteins B and C. Reprod. Biol. Endocrinol. **7**:84.
374. **Geisler, W. M., R. J. Suchland, D. D. Rockey, and W. E. Stamm**. 2001. Epidemiology and clinical manifestations of unique *Chlamydia trachomatis* isolates that occupy nonfusogenic inclusions. J. Infect. Dis. **184**:879-884.
375. **Grotenbreg, G. M., N. R. Roan, E. Guillen, R. Meijers, J. H. Wang, G. W. Bell, M. N. Starnbach, and H. L. Ploegh**. 2008. Discovery of CD8+ T cell epitopes in *Chlamydia trachomatis* infection through use of caged class I MHC tetramers. Proc. Natl. Acad. Sci. U. S. A **105**:3831-3836.
376. **Jeffrey, B. M., R. J. Suchland, S. G. Eriksen, K. M. Sandoz, and D. D. Rockey**. 2013. Genomic and phenotypic characterization of *in vitro*-generated *Chlamydia trachomatis* recombinants. BMC. Microbiol. **13**:142.
377. **Kuo, C. C. and T. Grayston**. 1976. Interaction of *Chlamydia trachomatis* organisms and HeLa 229 cells. Infect. Immun. **13**:1103-1109.
378. **Molleken, K., E. Schmidt, and J. H. Hegemann**. 2010. Members of the Pmp protein family of *Chlamydia pneumoniae* mediate adhesion to human cells via short repetitive peptide motifs. Mol. Microbiol. **78**:1004-1017.
379. **Carlson, J. H., W. M. Whitmire, D. D. Crane, L. Wicke, K. Virtaneva, D. E. Sturdevant, J. J. Kupko, III, S. F. Porcella, N. Martinez-Orengo, R. A. Heinzen, L. Kari, and H. D. Caldwell**. 2008. The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect. Immun. **76**:2273-2283.
380. **Gong, S., Z. Yang, L. Lei, L. Shen, and G. Zhong**. 2013. Characterization of *Chlamydia trachomatis* plasmid-encoded open reading frames. J. Bacteriol. **195**:3819-3826.
381. **Ramsey, K. H., J. H. Schripsema, B. J. Smith, Y. Wang, B. C. Jham, K. P. O'Hagan, N. R. Thomson, A. K. Murthy, R. J. Skilton, P. Chu, and I. N. Clarke**. 2014. Plasmid CDS5 influences infectivity and virulence in a mouse model of *Chlamydia trachomatis* urogenital infection. Infect. Immun. **82**:3341-3349.
382. **Wang, Y., L. T. Cutcliffe, R. J. Skilton, K. Persson, C. Bjartling, and I. N. Clarke**. 2013. Transformation of a plasmid-free, genital tract isolate of *Chlamydia trachomatis* with a plasmid vector carrying a deletion in CDS6 revealed that this gene regulates inclusion phenotype. Pathog. Dis. **67**:100-103.
383. **O'Connell, C. M., Y. M. AbdelRahman, E. Green, H. K. Darville, K. Saira, B. Smith, T. Darville, A. M. Scurlock, C. R. Meyer, and R. J. Belland**. 2011. Toll-like receptor 2 activation by *Chlamydia trachomatis* is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by C. *trachomatis* but not by C. *muridarum*. Infect. Immun. **79**:1044-1056.
384. **Song, L., J. H. Carlson, B. Zhou, K. Virtaneva, W. M. Whitmire, G. L. Sturdevant, S. F. Porcella, G. McClarty, and H. D. Caldwell**. 2014. Plasmid-mediated transformation tropism of chlamydial biovars. Pathog. Dis. **70**:189-193.

385. **Wang, Y., L. T. Cutcliffe, R. J. Skilton, K. H. Ramsey, N. R. Thomson, and I. N. Clarke**. 2014. The genetic basis of plasmid tropism between *Chlamydia trachomatis* and *Chlamydia muridarum*. Pathog. Dis. **72**:19-23.
386. **Comanducci, M., S. Ricci, R. Cevenini, and G. Ratti**. 1990. Diversity of the *Chlamydia trachomatis* common plasmid in biovars with different pathogenicity. Plasmid **23**:149-154.
387. **Thomas, N. S., M. Lusher, C. C. Storey, and I. N. Clarke**. 1997. Plasmid diversity in *Chlamydia*. Microbiology **143 (Pt 6)**:1847-1854.
388. **Christerson, L., H. J. de Vries, B. B. de, C. A. Gaydos, B. Henrich, S. Hoffmann, J. Schachter, J. Thorvaldsen, M. Vall-Mayans, M. Klint, B. Herrmann, and S. A. Morre**. 2010. Typing of lymphogranuloma venereum *Chlamydia trachomatis* strains. Emerg. Infect. Dis. **16**:1777-1779.
389. **Herida, M., B. B. de, P. Sednaoui, C. Scieux, N. Lemarchand, G. Kreplak, M. Clerc, J. Timsit, V. Goulet, J. C. Desenclos, and C. Semaille**. 2006. Rectal lymphogranuloma venereum surveillance in France 2004-2005. Euro. Surveill **11**:155-156.
390. **Spaargaren, J., J. Schachter, J. Moncada, H. J. de Vries, H. S. Fennema, A. S. Pena, R. A. Coutinho, and S. A. Morre**. 2005. Slow epidemic of lymphogranuloma venereum L2b strain. Emerg. Infect. Dis. **11**:1787-1788.
391. **Kang, E. X., X. Gao, Y. P. Yin, F. S. Wang, W. D. Yao, X. Q. Gong, and X. S. Chen**. 2007. Lymphogranuloma venereum caused by *Chlamydia trachomatis* serovar L3: a case report. Chin Med. J. (Engl.) **120**:601-604.
392. **Gogol, E. B., C. A. Cummings, R. C. Burns, and D. A. Relman**. 2007. Phase variation and microevolution at homopolymeric tracts in Bordetella pertussis. BMC. Genomics **8**:122.
393. **Orsi, R. H., B. M. Bowen, and M. Wiedmann**. 2010. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. BMC. Genomics **11**:102.
394. **Salaun, L., L. A. Snyder, and N. J. Saunders**. 2003. Adaptation by phase variation in pathogenic bacteria. Adv. Appl. Microbiol. **52**:263-301.
395. **van der Woude, M. W. and A. J. Baumler**. 2004. Phase and antigenic variation in bacteria. Clin. Microbiol. Rev. **17**:581-611, table.
396. **Grimwood, J., L. Olinger, and R. S. Stephens**. 2001. Expression of *Chlamydia pneumoniae* polymorphic membrane protein family genes. Infect. Immun. **69**:2383-2389.
397. **Pedersen, A. S., G. Christiansen, and S. Birkelund**. 2001. Differential expression of Pmp10 in cell culture infected with *Chlamydia pneumoniae* CWL029. FEMS Microbiol. Lett. **203**:153-159.
398. **Graham, J. E. and J. E. Clark-Curtiss**. 1999. Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). Proc. Natl. Acad. Sci. U. S. A **96**:11554-11559.
399. **Verweij, S. P., A. Catsburg, S. Ouburg, A. Lombardi, R. Heijmans, F. Dutly, R. Frei, S. A. Morre, and D. Goldenberger**. 2011. Lymphogranuloma venereum variant L2b-specific polymerase chain reaction: insertion used to close an epidemiological gap. Clin. Microbiol. Infect. **17**:1727-1730.
400. **Schachter, J. and J. Moncada**. 2005. Lymphogranuloma venereum: how to turn an endemic disease into an outbreak of a new disease? Start looking. Sex Transm. Dis. **32**:331-332.
401. **Fenton, K. A. and J. Imrie**. 2005. Increasing rates of sexually transmitted diseases in homosexual men in Western europe and the United States: why? Infect. Dis. Clin. North Am. **19**:311-331.
402. **Korhonen, S., E. Hiltunen-Back, and M. Puolakkainen**. 2012. Genotyping of *Chlamydia trachomatis* in rectal and pharyngeal specimens: identification of LGV genotypes in Finland. Sex Transm. Infect. **88**:465-469.
403. **Sethi, G., E. Iason-Jones, J. Richens, N. T. Annan, D. Hawkins, A. Ekbote, S. Alexander, and J. White**. 2009. Lymphogranuloma venereum presenting as genital ulceration and inguinal syndrome in men who have sex with men in London, UK. Sex Transm. Infect. **85**:165-170.
404. **Stary, G. and A. Stary**. 2008. Lymphogranuloma venereum outbreak in Europe. J. Dtsch. Dermatol. Ges. **6**:935-940.
405. **Vall-Mayans, M., J. Isaksson, E. Caballero, B. Salles, and B. Herrmann**. 2014. Bubonic lymphogranuloma venereum with multidrug treatment failure. Int. J. STD AIDS **25**:306-308.
406. **Verma, A. and A. T. Maurelli**. 2003. Identification of two eukaryote-like serine/threonine kinases encoded by *Chlamydia trachomatis* serovar L2 and characterization of interacting partners of Pkn1. Infect. Immun. **71**:5772-5784.

407. **Binet, R., R. E. Fernandez, D. J. Fisher, and A. T. Maurelli**. 2011. Identification and characterization of the *Chlamydia trachomatis* L2 S-adenosylmethionine transporter. MBio. **2**:e00051-11.
408. **Chen, D., L. Lei, C. Lu, R. Flores, M. P. DeLisa, T. C. Roberts, F. E. Romesberg, and G. Zhong**. 2010. Secretion of the chlamydial virulence factor CPAF requires the Sec-dependent pathway. Microbiology **156**:3031-3040.
409. **Kendall, B. A., K. D. Tardif, and R. Schlaberg**. 2014. *Chlamydia trachomatis* L serovars and dominance of novel L2b *ompA* variants, U.S.A. Sex Transm. Infect. **90**:336.
410. **Luo, H., J. Tang, R. Friedman, and A. L. Hughes**. 2011. Ongoing purifying selection on intergenic spacers in group A streptococcus. Infect. Genet. Evol. **11**:343-348.
411. **Perez, N., J. Trevino, Z. Liu, S. C. Ho, P. Babitzke, and P. Sumby**. 2009. A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus. PLoS. One. **4**:e7668.
412. **Tsai, C. H., R. Liao, B. Chou, M. Palumbo, and L. M. Contreras**. 2015. Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. J. Bacteriol. **197**:40-50.
413. **Tsoy, O. V., M. A. Pyatnitskiy, M. D. Kazanov, and M. S. Gelfand**. 2012. Evolution of transcriptional regulation in closely related bacteria. BMC. Evol. Biol. **12**:200.
414. **Bao, X., N. D. Pachikara, C. B. Oey, A. Balakrishnan, L. F. Westblade, M. Tan, T. Chase, Jr., B. E. Nickels, and H. Fan**. 2011. Non-coding nucleotides and amino acids near the active site regulate peptide deformylase expression and inhibitor susceptibility in *Chlamydia trachomatis*. Microbiology **157**:2569-2581.
415. **Gronewold, T. M. and D. Kaiser**. 2007. Mutations of the act promoter in Myxococcus xanthus. J. Bacteriol. **189**:1836-1844.
416. **Rocha, E. P., J. M. Smith, L. D. Hurst, M. T. Holden, J. E. Cooper, N. H. Smith, and E. J. Feil**. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. J. Theor. Biol. **239**:226-235.
417. **Akers, J. C. and M. Tan**. 2006. Molecular mechanism of tryptophan-dependent transcriptional regulation in *Chlamydia trachomatis*. J. Bacteriol. **188**:4236-4243.
418. **Carlson, J. H., H. Wood, C. Roshick, H. D. Caldwell, and G. McClarty**. 2006. *In vivo* and *in vitro* studies of *Chlamydia trachomatis* TrpR:DNA interactions. Mol. Microbiol. **59**:1678-1691.
419. **Sniegowski, P. D., P. J. Gerrish, T. Johnson, and A. Shaver**. 2000. The evolution of mutation rates: separating causes from consequences. Bioessays **22**:1057-1066.
420. **Conrad, T. M., N. E. Lewis, and B. O. Palsson**. 2011. Microbial laboratory evolution in the era of genome-scale science. Mol. Syst. Biol. **7**:509.
421. **Labiran, C., I. N. Clarke, L. T. Cutcliffe, Y. Wang, R. J. Skilton, K. Persson, C. Bjartling, B. Herrmann, L. Christerson, and P. Marsh**. 2012. Genotyping markers used for multi locus VNTR analysis with *ompA* (MLVA-*ompA*) and multi sequence typing (MST) retain stability in *Chlamydia trachomatis*. Front Cell Infect. Microbiol. **2**:68.
422. **Stothard, D. R., P. B. Van Der, N. J. Smith, and R. B. Jones**. 1998. Effect of serial passage in tissue culture on sequence of omp1 from *Chlamydia trachomatis* clinical isolates. J. Clin. Microbiol. **36**:3686-3688.
423. **Van Valen L**. 1973. A new evolutionary law., *In* . Evol. Theory, 1, 1-30.
424. **Lambden, P. R., M. A. Pickett, and I. N. Clarke**. 2006. The effect of penicillin on *Chlamydia trachomatis* DNA replication. Microbiology **152**:2573-2578.
425. **Gomes, J. P.** 2012. Genomic stability of genotyping markers in *Chlamydia trachomatis*. Front Cell Infect. Microbiol. **2**:77.
426. **Schachter, J. and K. F. Meyer**. 1969. Lymphogranuloma venereum. II. Characterization of some recently isolated strains. J. Bacteriol. **99**:636-638.
427. **Catry, M. A., M. J. Borrego, J. Cardoso, J. Azevedo, and I. Santo**. 1995. Comparison of the Amplicor *Chlamydia trachomatis* test and cell culture for the detection of urogenital chlamydial infections. Genitourin. Med. **71**:247-250.
428. **Caldwell, H. D., J. Kromhout, and J. Schachter**. 1981. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. Infect. Immun. **31**:1161-1176.
429. **Lefebvre, J. F. and J. Orfila**. 1980. [Polypeptidic composition analysis of "*Chlamydia*" purified elementary bodies by polyacrylamide slab gel electrophoresis (author's transl)]. Ann. Microbiol. (Paris) **131B**:223-237.

430. **Gerrish, P. J. and R. E. Lenski**. 1998. The fate of competing beneficial mutations in an asexual population. Genetica **102-103**:127-144.
431. **Russell, M., T. Darville, K. Chandra-Kuntal, B. Smith, C. W. Andrews, Jr., and C. M. O'Connell**. 2011. Infectivity acts as *in vivo* selection for maintenance of the chlamydial cryptic plasmid. Infect. Immun. **79**:98-107.
432. **Suchland, R. J., B. M. Jeffrey, M. Xia, A. Bhatia, H. G. Chu, D. D. Rockey, and W. E. Stamm**. 2008. Identification of concomitant infection with *Chlamydia trachomatis* IncA-negative mutant and wild-type strains by genomic, transcriptional, and biological characterizations. Infect. Immun. **76**:5438-5446.
433. **Mathews, S. A., K. M. Volp, and P. Timms**. 1999. Development of a quantitative gene expression assay for *Chlamydia trachomatis* identified temporal expression of sigma factors. FEBS Lett. **458**:354-358.
434. **Miyairi, I., O. S. Mahdi, S. P. Ouellette, R. J. Belland, and G. I. Byrne**. 2006. Different growth rates of *Chlamydia trachomatis* biovars reflect pathotype. J. Infect. Dis. **194**:350-357.
435. **Iliopoulos, I., S. Tsoka, M. A. Andrade, A. J. Enright, M. Carroll, P. Poullet, V. Promponas, T. Liakopoulos, G. Palaios, C. Pasquier, S. Hamodrakas, J. Tamames, A. T. Yagnik, A. Tramontano, D. Devos, C. Blaschke, A. Valencia, D. Brett, D. Martin, C. Leroy, I. Rigoutsos, C. Sander, and C. A. Ouzounis**. 2003. Evaluation of annotation strategies using an entire genome sequence. Bioinformatics. **19**:717-726.
436. **Belland, R. J., M. A. Scidmore, D. D. Crane, D. M. Hogan, W. Whitmire, G. McClarty, and H. D. Caldwell**. 2001. *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proc. Natl. Acad. Sci. U. S. A **98**:13984-13989.
437. **Herring, C. D., A. Raghunathan, C. Honisch, T. Patel, M. K. Applebee, A. R. Joyce, T. J. Albert, F. R. Blattner, B. D. van den, C. R. Cantor, and B. O. Palsson**. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat. Genet. **38**:1406-1412.
438. **Bush, R. M., C. B. Smith, N. J. Cox, and W. M. Fitch**. 2000. Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution. Proc. Natl. Acad. Sci. U. S. A **97**:6974-6980.
439. **Sullivan, N., Y. Sun, Q. Sattentau, M. Thali, D. Wu, G. Denisova, J. Gershoni, J. Robinson, J. Moore, and J. Sodroski**. 1998. CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. J. Virol. **72**:4694-4703.
440. **Martinez, M. A., N. Verdaguer, M. G. Mateu, and E. Domingo**. 1997. Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. Proc. Natl. Acad. Sci. U. S. A **94**:6798-6802.
441. **Binet, R. and A. T. Maurelli**. 2005. Frequency of spontaneous mutations that confer antibiotic resistance in *Chlamydia* spp. Antimicrob. Agents Chemother. **49**:2865-2873.
442. **Kutlin, A., S. Kohlhoff, P. Roblin, M. R. Hammerschlag, and P. Riska**. 2005. Emergence of resistance to rifampin and rifalazil in *Chlamydophila pneumoniae* and *Chlamydia trachomatis*. Antimicrob. Agents Chemother. **49**:903-907.
443. **Rupp, J., W. Solbach, and J. Gieffers**. 2008. Variation in the mutation frequency determining quinolone resistance in *Chlamydia trachomatis* serovars L2 and D. J. Antimicrob. Chemother. **61**:91-94.
444. **Sandoz, K. M., S. G. Eriksen, B. M. Jeffrey, R. J. Suchland, T. E. Putman, D. E. Hruby, R. Jordan, and D. D. Rockey**. 2012. Resistance to a novel antichlamydial compound is mediated through mutations in *Chlamydia trachomatis* secY. Antimicrob. Agents Chemother. **56**:4296-4302.
445. **Suchland, R. J., A. Bourillon, E. Denamur, W. E. Stamm, and D. M. Rothstein**. 2005. Rifampin-resistant RNA polymerase mutants of *Chlamydia trachomatis* remain susceptible to the ansamycin rifalazil. Antimicrob. Agents Chemother. **49**:1120-1126.
446. **Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow**. 1998. Rates of spontaneous mutation. Genetics **148**:1667-1686.
447. **Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim**. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli.* Nature **461**:1243-1247.
448. **Dettman, J. R., N. Rodrigue, A. H. Melnyk, A. Wong, S. F. Bailey, and R. Kassen**. 2012. Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. Mol. Ecol. **21**:2058-2077.

449. **Meyer, Y., B. B. Buchanan, F. Vignols, and J. P. Reichheld**. 2009. Thioredoxins and glutaredoxins: unifying elements in redox biology. Annu. Rev. Genet. **43**:335-367.
450. **Cooper, T. F., D. E. Rozen, and R. E. Lenski**. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A **100**:1072-1077.
451. **Elena, S. F. and R. E. Lenski**. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. **4**:457-469.
452. **Lee, D. H. and B. O. Palsson**. 2010. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. Appl. Environ. Microbiol. **76**:4158-4168.
453. **Zeyl, C.** 2005. The number of mutations selected during adaptation in a laboratory population of *Saccharomyces cerevisiae*. Genetics **169**:1825-1831.
454. **Taubenberger, J. K., A. H. Reid, R. M. Lourens, R. Wang, G. Jin, and T. G. Fanning**. 2005. Characterization of the 1918 influenza virus polymerase genes. Nature **437**:889-893.
455. **O'Connell, C. M. and K. M. Nicks**. 2006. A plasmid-cured *Chlamydia muridarum* strain displays altered plaque morphology and reduced infectivity in cell culture. Microbiology **152**:1601-1607.
456. **Biskup, U. G., F. Strle, and E. Ruzic-Sabljic**. 2011. Loss of plasmids of Borrelia burgdorferi sensu lato during prolonged *in vitro* cultivation. Plasmid **66**:1-6.
457. **Fong, S. S., A. R. Joyce, and B. O. Palsson**. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Res. **15**:1365-1372.
458. **Vannucci, F. A., D. N. Foster, and C. J. Gebhart**. 2012. Comparative transcriptional analysis of homologous pathogenic and non-pathogenic *Lawsonia intracellularis* isolates in infected porcine cells. PLoS. One. **7**:e46708.
459. **Stoebel, D. M., K. Hokamp, M. S. Last, and C. J. Dorman**. 2009. Compensatory evolution of gene regulation in response to stress by *Escherichia coli* lacking RpoS. PLoS. Genet. **5**:e1000671.
460. **Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst**. 2010. Antagonistic coevolution accelerates molecular evolution. Nature **464**:275-278.
461. **Omsland, A., D. C. Cockrell, D. Howe, E. R. Fischer, K. Virtaneva, D. E. Sturdevant, S. F. Porcella, and R. A. Heinzen**. 2009. Host cell-free growth of the Q fever bacterium Coxiella burnetii. Proc. Natl. Acad. Sci. U. S. A **106**:4430-4434.
462. **Kersh, G. J., L. D. Oliver, J. S. Self, K. A. Fitzpatrick, and R. F. Massung**. 2011. Virulence of pathogenic Coxiella burnetii strains after growth in the absence of host cells. Vector. Borne. Zoonotic. Dis. **11**:1433-1438.
463. **Barrick, J. E. and R. E. Lenski**. 2013. Genome dynamics during experimental evolution. Nat. Rev. Genet. **14**:827-839.
464. **Conrad, T. M., A. R. Joyce, M. K. Applebee, C. L. Barrett, B. Xie, Y. Gao, and B. O. Palsson**. 2009. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. **10**:R118.
465. **Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski**. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J. Bacteriol. **183**:2834-2841.
466. **Friedman, L., J. D. Alder, and J. A. Silverman**. 2006. Genetic changes that correlate with reduced susceptibility to daptomycin in Staphylococcus aureus. Antimicrob. Agents Chemother. **50**:2137-2145.
467. **Jansen, G., C. Barbosa, and H. Schulenburg**. 2013. Experimental evolution as an efficient tool to dissect adaptive paths to antibiotic resistance. Drug Resist. Updat. **16**:96-107.
468. **Zinser, E. R., D. Schneider, M. Blot, and R. Kolter**. 2003. Bacterial evolution through the selective loss of beneficial Genes. Trade-offs in expression involving two loci. Genetics **164**:1271-1277.
469. **Domenech, P. and M. B. Reed**. 2009. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown *in vitro*: implications for virulence studies. Microbiology **155**:3532-3543.
470. **Somerville, G. A., S. B. Beres, J. R. Fitzgerald, F. R. DeLeo, R. L. Cole, J. S. Hoff, and J. M. Musser**. 2002. *In vitro* serial passage of *Staphylococcus aureus*: changes in

physiology, virulence factor production, and agr nucleotide sequence. J. Bacteriol. **184**:1430-1437.
471. **Velicer, G. J., G. Raddatz, H. Keller, S. Deiss, C. Lanz, I. Dinkelacker, and S. C. Schuster**. 2006. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. Proc. Natl. Acad. Sci. U. S. A **103**:8107-8112.
472. **Chen, C., Z. Zhou, T. Conrad, Z. Yang, J. Dai, Z. Li, Y. Wu, and G. Zhong**. 2015. *In Vitro* Passage Selects for *Chlamydia muridarum* with Enhanced Infectivity in Cultured Cells but Attenuated Pathogenicity in Mouse Upper Genital Tract. Infect. Immun. **83**:1881-1892.
473. **Borges, V. and J. P. Gomes**. 2015. Deep comparative genomics among *Chlamydia trachomatis* lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation. Infect. Genet. Evol. **32**:74-88.
474. **Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov**. 2011. Integrative genomics viewer. Nat. Biotechnol. **29**:24-26.
475. **Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen**. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A **99**:9697-9702.
476. **Engstrom, P., L. Bailey, T. Onskog, S. Bergstrom, and J. Johansson**. 2010. A comparative study of RNA and DNA as internal gene expression controls early in the developmental cycle of *Chlamydia pneumoniae*. FEMS Immunol. Med. Microbiol. **58**:244-253.
477. **Selinger, D. W., R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow**. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res. **13**:216-223.
478. **Anders, S. and W. Huber**. 2010. Differential expression analysis for sequence count data. Genome Biol. **11**:R106.
479. **Benjamini Y and Hochberg Y**. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, p. 289-300. *In* . Journal of the Royal Statistical Society.
480. **Borges, V., M. Pinheiro, L. Vieira, D. A. Sampaio, A. Nunes, M. J. Borrego, and J. P. Gomes**. 2014. Complete Genome Sequence of *Chlamydia trachomatis* Ocular Serovar C Strain TW-3. Genome Announc. **2**.
481. **Lee, C. J., X. Liang, X. Chen, D. Zeng, S. H. Joo, H. S. Chung, A. W. Barb, S. M. Swanson, R. A. Nicholas, Y. Li, E. J. Toone, C. R. Raetz, and P. Zhou**. 2011. Species-specific and inhibitor-dependent conformations of LpxC: implications for antibiotic design. Chem. Biol. **18**:38-47.
482. **Nguyen, B. D., D. Cunningham, X. Liang, X. Chen, E. J. Toone, C. R. Raetz, P. Zhou, and R. H. Valdivia**. 2011. Lipooligosaccharide is required for the generation of infectious elementary bodies in *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S. A **108**:10284-10289.
483. **Engstrom, P., B. D. Nguyen, J. Normark, I. Nilsson, R. J. Bastidas, A. Gylfe, M. Elofsson, K. A. Fields, R. H. Valdivia, H. Wolf-Watz, and S. Bergstrom**. 2013. Mutations in *hemG* mediate resistance to salicylidene acylhydrazides, demonstrating a novel link between protoporphyrinogen oxidase (HemG) and *Chlamydia trachomatis* infectivity. J. Bacteriol. **195**:4221-4230.
484. **von, E. C., C. Heilmann, R. A. Proctor, C. Woltz, G. Peters, and F. Gotz**. 1997. A site-directed Staphylococcus aureus hemB mutant is a small-colony variant which persists intracellularly. J. Bacteriol. **179**:4706-4712.
485. **Brinkworth, A. J., D. S. Malcolm, A. T. Pedrosa, K. Roguska, S. Shahbazian, J. E. Graham, R. D. Hayward, and R. A. Carabeo**. 2011. *Chlamydia trachomatis* Slc1 is a type III secretion chaperone that enhances the translocation of its invasion effector substrate TARP. Mol. Microbiol. **82**:131-144.
486. **Betts-Hampikian, H. J. and K. A. Fields**. 2010. The chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. Front Microbiol. **1**:114.
487. **Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. J. Pallen**. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. **30**:434-439.
488. **Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu**. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC. Genomics **13**:341.

489. **Jerome, J. P., J. A. Bell, A. E. Plovanich-Jones, J. E. Barrick, C. T. Brown, and L. S. Mansfield**. 2011. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. PLoS. One. **6**:e16399.
490. **Weisberg, R. A. and S. Adhya**. 1977. Illegitimate recombination in bacteria and bacteriophage. Annu. Rev. Genet. **11**:451-473.
491. **Michel, B.** 2000. Replication fork arrest and DNA recombination. Trends Biochem. Sci. **25**:173-178.
492. **Lovett, S. T.** 2004. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol. Microbiol. **52**:1243-1253.
493. **Andersson, S. G. and C. G. Kurland**. 1998. Reductive evolution of resident genomes. Trends Microbiol. **6**:263-268.
494. **Rocha, E. P.** 2003. An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: from duplications to genome reduction. Genome Res. **13**:1123-1132.
495. **Bonner, C. A., G. I. Byrne, and R. A. Jensen**. 2014. *Chlamydia* exploit the mammalian tryptophan-depletion defense strategy as a counter-defensive cue to trigger a survival state of persistence. Front Cell Infect. Microbiol. **4**:17.
496. **Kabeya, Y., H. Nakanishi, K. Suzuki, T. Ichikawa, Y. Kondou, M. Matsui, and S. Y. Miyagishima**. 2010. The YlmG protein has a conserved function related to the distribution of nucleoids in chloroplasts and cyanobacteria. BMC. Plant Biol. **10**:57.
497. **Weiling, H., Y. Xiaowen, L. Chunmei, and X. Jianping**. 2013. Function and evolution of ubiquitous bacterial signaling adapter phosphopeptide recognition domain FHA. Cell Signal. **25**:660-665.
498. **Subtil, A., A. Collingro, and M. Horn**. 2014. Tracing the primordial *Chlamydiae*: extinct parasites of plants? Trends Plant Sci. **19**:36-43.
499. **Alifano, P., C. B. Bruni, and M. S. Carlomagno**. 1994. Control of mRNA processing and decay in prokaryotes. Genetica **94**:157-172.
500. **Deana, A. and J. G. Belasco**. 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev. **19**:2526-2533.
501. **Fields, K. A., E. R. Fischer, D. J. Mead, and T. Hackstadt**. 2005. Analysis of putative *Chlamydia trachomatis* chaperones Scc2 and Scc3 and their use in the identification of type III secretion substrates. J. Bacteriol. **187**:6466-6478.
502. **Hou, S., L. Lei, Z. Yang, M. Qi, Q. Liu, and G. Zhong**. 2013. *Chlamydia trachomatis* outer membrane complex protein B (OmcB) is processed by the protease CPAF. J. Bacteriol. **195**:951-957.
503. **Jorgensen, I., M. M. Bednar, V. Amin, B. K. Davis, J. P. Ting, D. G. McCafferty, and R. H. Valdivia**. 2011. The *Chlamydia* protease CPAF regulates host and bacterial proteins to maintain pathogen vacuole integrity and promote virulence. Cell Host. Microbe **10**:21-32.
504. **Lad, S. P., J. Li, C. J. da Silva, Q. Pan, S. Gadwal, R. J. Ulevitch, and E. Li**. 2007. Cleavage of p65/RelA of the NF-kappaB pathway by *Chlamydia*. Proc. Natl. Acad. Sci. U. S. A **104**:2933-2938.
505. **Fahr, M. J., A. L. Douglas, W. Xia, and T. P. Hatch**. 1995. Characterization of late gene promoters of *Chlamydia trachomatis*. J. Bacteriol. **177**:4252-4260.
506. **Griffiths, E., M. S. Ventresca, and R. S. Gupta**. 2006. BLAST screening of chlamydial genomes to identify signature proteins that are unique for the *Chlamydiales, Chlamydia*ceae, *Chlamydophila* and *Chlamydia* groups of species. BMC. Genomics **7**:14.
507. **Birkelund, S., M. Morgan-Fisher, E. Timmerman, K. Gevaert, A. C. Shaw, and G. Christiansen**. 2009. Analysis of proteins in *Chlamydia trachomatis* L2 outer membrane complex, COMC. FEMS Immunol. Med. Microbiol. **55**:187-195.
508. **Yeruva, L., G. S. Myers, N. Spencer, H. H. Creasy, N. E. Adams, A. T. Maurelli, G. R. McChesney, M. A. Cleves, J. Ravel, A. Bowlin, and R. G. Rank**. 2014. Early microRNA expression profile as a prognostic biomarker for the development of pelvic inflammatory disease in a mouse model of chlamydial genital infection. MBio. **5**:e01241-14.
509. **Wang, Y., S. Kahane, L. T. Cutcliffe, R. J. Skilton, P. R. Lambden, and I. N. Clarke**. 2011. Development of a transformation system for *Chlamydia trachomatis*: restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. PLoS. Pathog. **7**:e1002258.
510. **Ding, H., S. Gong, Y. Tian, Z. Yang, R. Brunham, and G. Zhong**. 2013. Transformation of sexually transmitted infection-causing serovars of *Chlamydia trachomatis* using Blasticidin for selection. PLoS. One. **8**:e80534.

511. **O'Neill, C. E., H. M. Seth-Smith, P. B. Van Der, S. R. Harris, N. R. Thomson, L. T. Cutcliffe, and I. N. Clarke**. 2013. *Chlamydia trachomatis* clinical isolates identified as tetracycline resistant do not exhibit resistance *in vitro*: whole-genome sequencing reveals a mutation in *porB* but no evidence for tetracycline resistance genes. Microbiology **159**:748-756.
512. **Engstrom, P., K. S. Krishnan, B. D. Ngyuen, E. Chorell, J. Normark, J. Silver, R. J. Bastidas, M. D. Welch, S. J. Hultgren, H. Wolf-Watz, R. H. Valdivia, F. Almqvist, and S. Bergstrom**. 2015. A 2-pyridone-amide inhibitor targets the glucose metabolism pathway of *Chlamydia trachomatis*. MBio. **6**:e02304-e02314.
513. **Storz, G., J. Vogel, and K. M. Wassarman**. 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell **43**:880-891.
514. **Ortega, A. D., J. J. Quereda, M. G. Pucciarelli, and P. F. Garcia-del**. 2014. Non-coding RNA regulation in pathogenic bacteria located inside eukaryotic cells. Front Cell Infect. Microbiol. **4**:162.
515. **Cooper, K. K., M. A. Cooper, A. Zuccolo, and L. A. Joens**. 2013. Re-sequencing of a virulent strain of *Campylobacter jejuni* NCTC11168 reveals potential virulence factors. Res. Microbiol. **164**:6-11.
516. **Bao, X., A. Gylfe, G. L. Sturdevant, Z. Gong, S. Xu, H. D. Caldwell, M. Elofsson, and H. Fan**. 2014. Benzylidene acylhydrazides inhibit chlamydial growth in a type III secretion- and iron chelation-independent manner. J. Bacteriol. **196**:2989-3001.
517. **Kari, L., W. M. Whitmire, J. H. Carlson, D. D. Crane, N. Reveneau, D. E. Nelson, D. C. Mabey, R. L. Bailey, M. J. Holland, G. McClarty, and H. D. Caldwell**. 2008. Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. J. Infect. Dis. **197**:449-456.
518. **Viratyosin, W., L. A. Campbell, C. C. Kuo, and D. D. Rockey**. 2002. Intrastrain and interstrain genetic variation within a paralogous gene family in *Chlamydia pneumoniae*. BMC. Microbiol. **2**:38.
519. **Nelson, D. E., D. P. Virok, H. Wood, C. Roshick, R. M. Johnson, W. M. Whitmire, D. D. Crane, O. Steele-Mortimer, L. Kari, G. McClarty, and H. D. Caldwell**. 2005. Chlamydial IFN-gamma immune evasion is linked to host infection tropism. Proc. Natl. Acad. Sci. U. S. A **102**:10658-10663.
520. **Nelson, D. E., L. D. Taylor, J. G. Shannon, W. M. Whitmire, D. D. Crane, G. McClarty, H. Su, L. Kari, and H. D. Caldwell**. 2007. Phenotypic rescue of *Chlamydia trachomatis* growth in IFN-gamma treated mouse cells by irradiated *Chlamydia muridarum*. Cell Microbiol. **9**:2289-2298.
521. **Al-Zeer, M. A., H. M. Al-Younes, D. Lauster, L. M. Abu, and T. F. Meyer**. 2013. Autophagy restricts *Chlamydia trachomatis* growth in human macrophages via IFNG-inducible guanylate binding proteins. Autophagy. **9**:50-62.
522. **Tietzel, I., C. El-Haibi, and R. A. Carabeo**. 2009. Human guanylate binding proteins potentiate the anti-*Chlamydia* effects of interferon-gamma. PLoS. One. **4**:e6499.
523. **Lukacova, M., M. Baumann, L. Brade, U. Mamat, and H. Brade**. 1994. Lipopolysaccharide smooth-rough phase variation in bacteria of the genus *Chlamydia*. Infect. Immun. **62**:2270-2276.
524. **Moxon, R., C. Bayliss, and D. Hood**. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu. Rev. Genet. **40**:307-333.
525. **Cowley, S. C., S. V. Myltseva, and F. E. Nano**. 1996. Phase variation in Francisella tularensis affecting intracellular growth, lipopolysaccharide antigenicity and nitric oxide production. Mol. Microbiol. **20**:867-874.
526. **Darwin C**. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. (Murray, London).

Supplemental material

Supplemental Table S2.1. List of strains studied.

Strain name	GenBank Accession number	Reference or source
Chlamydia trachomatis A/2497	NC_016798	(37)
Chlamydia trachomatis A/363	HE601796	(37)
Chlamydia trachomatis A/5291	HE601810	(37)
Chlamydia trachomatis A/7249	HE601797	(37)
Chlamydia trachomatis A/Har13	NC_007429	(165)
Chlamydia trachomatis B/TZ1A828/OT	NC_012687	(166)
Chlamydia trachomatis B/Jali20/OT	NC_012686	(166)
Chlamydia trachomatis Ba/Apache-2	---[a]	This publication
Chlamydia trachomatis C/TW3	---[a]	This publication
Chlamydia trachomatis D/SotonD1	HE601798	(37)
Chlamydia trachomatis D/SotonD5	HE601799	(37)
Chlamydia trachomatis D/SotonD6	HE601800	(37)
Chlamydia trachomatis D/UW3	NC_000117	(115)
Chlamydia trachomatis E/SW2	FN652779	(284)
Chlamydia trachomatis E/SW3	HE601801	(37)
Chlamydia trachomatis E/SotonE4	HE601802	(37)
Chlamydia trachomatis E/SotonE8	HE601803	(37)
Chlamydia trachomatis E/11023	NC_017431	(172)
Chlamydia trachomatis E/150	NC_017439	(172)
Chlamydia trachomatis E/Bour	HE601870	(37)
Chlamydia trachomatis F/SW4	HE601804	(37)
Chlamydia trachomatis F/SW5	HE601805	(37)
Chlamydia trachomatis F/SotonF3	HE601806	(37)
Chlamydia trachomatis F/IC-Cal3	---[a]	This publication
Chlamydia trachomatis G/9301	NC_017432	(172)
Chlamydia trachomatis G/9768	NC_017429	(172)
Chlamydia trachomatis G/11074	NC_017440	(172)
Chlamydia trachomatis G/11222	NC_017430	(172)
Chlamydia trachomatis G/SotonG1	HE601807	(37)
Chlamydia trachomatis G/UW57	---[a]	This publication
Chlamydia trachomatis H/UW43	---[a]	This publication
Chlamydia trachomatis I/UW12	---[a]	This publication
Chlamydia trachomatis Ia/SotonIa1	HE601808	(37)
Chlamydia trachomatis Ia/SotonIa3	HE601809	(37)
Chlamydia trachomatis J/6276	ABYD01000001	(172)
Chlamydia trachomatis J/UW36	---[a]	This publication
Chlamydia trachomatis K/SotonK1	HE601794	(37)
Chlamydia trachomatis K/UW31	---[a]	This publication
Chlamydia trachomatis L1/1322/p2	HE601951	(37)
Chlamydia trachomatis L1/115	HE601952	(37)
Chlamydia trachomatis L1/224	HE601953	(37)
Chlamydia trachomatis L1/440/LN	HE601950	(37)

Chlamydia trachomatis L2/25667R	HE601954	(37)
Chlamydia trachomatis L2/434/Bu	NC_010287	(46)
Chlamydia trachomatis L2b/UCH-1	NC_010280	(46)
Chlamydia trachomatis L2b/8200/07	HE601795	(37)
Chlamydia trachomatis L2b/UCH-2	HE601956	(37)
Chlamydia trachomatis L2b/Canada1	HE601963	(37)
Chlamydia trachomatis L2b/Canada2	HE601957	(37)
Chlamydia trachomatis L2b/LST	HE601958	(37)
Chlamydia trachomatis L2b/CV204	HE601960	(37)
Chlamydia trachomatis L2b/795	HE601949	(37)
Chlamydia trachomatis L2b/Ams1	HE601959	(37)
Chlamydia trachomatis L2b/Ams2	HE601961	(37)
Chlamydia trachomatis L2b/Ams3	HE601962.1	(37)
Chlamydia trachomatis L2b/Ams4	HE601964.1	(37)
Chlamydia trachomatis L2b/Ams5	HE601965.1	(37)
Chlamydia trachomatis L2c	NC_015744	(175)
Chlamydia trachomatis L3/404/LN	HE601955	(37)
Chlamydia muridarum Nigg	NC_002620	(29)

[a] As these historical prototype strains were not fully-sequenced so far, we amplified and sequenced the 75 genes under evaluation. The obtained nucleotide sequences have the GenBank accession numbers: JQ066324 - JQ066722.

Supplemental Table S2.2. Identification of genes and corresponding functional category.

ORF (gene)[a]	Functional category[b]	$d_N / d_S > 1$ [c]
CT048 (*yraL*)	HK	
CT049	other	
CT050	other	
CT051	other	
CT058	SEC	√
CT059 (*fer*)	HK	
CT105	other	√
CT115 (*incD*)	SEC	√
CT116 (*incE*)	SEC	√ *
CT117 (*incF*)	SEC	√
CT118 (*incG*)	SEC	√ *
CT119 (*incA*)	SEC	√
CT143	other	√
CT144	other	
CT147	SEC	√
CT192	SEC	
CT195	SEC	
CT214	SEC	√
CT222	SEC	√ *
CT223	SEC	√ *
CT228	SEC	√ *
CT229	SEC	√ *
CT232 (*incB*)	SEC	
CT233 (*incC*)	SEC	
CT249	SEC	√ *
CT288	SEC	√ *
CT293 (*accD*)	HK	
CT365	SEC	
CT383	SEC	√ *
CT412 (*pmpA*)	CEP	
CT413 (*pmpB*)	CEP	√ *
CT414 (*pmpC*)	CEP	√
CT442 (*crpA*)	SEC	√ *
CT443 (*omcB*)	CEP	√

CT456 (*tarp*)	SEC	√
CT529	SEC	√
CT618	SEC	
CT622	SEC	
CT623	other	
CT624 (*mviN*)	CEP	
CT653 (*yhbG*)	HK	
CT674 (*yscC*)	HK	
CT675 (*karG*)	HK	
CT676	other	
CT677 (*frr*)	HK	
CT678 (*pyrH*)	HK	
CT679 (*tsf*)	HK	
CT680 (*rs2* or *rpsB*)	HK	
CT681 (*ompA*)	CEP	
CT682 (*pbpB*)	CEP	
CT683 (*TPR motif-protein*)	other	
CT684	HK	
CT685	HK	
CT686 (*sufD*)	HK	
CT687 (*yfhO_1*)	HK	
CT688 (*parB*)	HK	
CT689 (*dppF*)	HK	
CT694	SEC	√ *
CT760 (*ftsW*)	HK	
CT783	other	
CT812 (*pmpD*)	CEP	
CT813	SEC	√ *
CT818 (*tyrP_2*)	HK	
CT852 (*yhgN*)	CEP	
CT859 (*ispH*)	HK	
CT860	SEC	
CT861	SEC	
CT862 (*lcrH_2*)	HK	
CT867	SEC	√ *
CT868	SEC	√ *
CT869 (*pmpE*)	CEP	
CT870 (*pmpF*)	CEP	
CT871 (*pmpG*)	CEP	
CT872 (*pmpH*)	CEP	
CT874 (*pmpI*)	CEP	

[a] Open reading frame (ORF) numbers are based on the D/UW3 strain genome annotation (GenBank accession number NC_000117).

[b] The 75 genes under evaluation were categorized according to their functional role: housekeeping genes (HKs), genes encoding well-known cell envelope proteins (CEPs), genes coding for proteins secreted into the cytosol of the host cells or to the inclusion membrane (SEC), and genes encoding proteins with unknown function or with non-consensual biological role (classified as "other").

[b] Genes exhibiting $d_N / d_S > 1$ in the analysis within *C. trachomatis* species (Figure 2.3). The symbol * refers to genes exhibiting both a d_N / d_S ratios above 1 and a significant *P* value (<0.05) in the codon-based Z-test of positive selection ($d_N/d_S > 1$).

Supplemental Table S2.3. Oligonucleotide primers used for PCR and sequencing.

ORF (gene)[a]	Primer[a]	Primer sequence (5' to 3')[a]	Primer location[a]	Amplicon size (bp)[a]
CT048 (*yraL*)	CT048-1[b]	GAGCCGGCTCTTTTAAATGGTTT	53104 - 53126	1059
	CT048-2[b]	GTCGACGGAACAGACGAAGAAA	54141 - 54162	
CT049[c]	---	---	---	---
CT050/CT051	CT050/51-1A[d, e]	TGGGCGCTGGTTATTAACTATTTG	377806 - 377829	3704
	CT050/51-2A[b, e]	GACCCCATCCCCTTTGGAGT	381490 - 381509	
	CT050/51-1D[b]	ACAAAGCGCTTTCAGAACATACAT	55452 - 55475	3589
	CT050/51-2D[b]	AGGGCGTCTTTTTCATGATTCTAT	59017 - 59040	
	CT050/51-1[f]	CTAAGAGTTATGTAGCTATC	55542 - 55561	
	CT050/51-1S[f]	AGTTAAGGGAGAGAATCTC	55593 - 55611	
	CT050/51-3[f]	TTGTAGTGTGCAAGATTGTC	55857 - 55876	
	CT050/51-4[f]	TTGTGCCACTACAATACCTT	58726 - 58745	
	CT050/51-5[f]	AACCTTTCCAATATCACCGT	56566 - 56585	
	CT050/51-6[f]	GCACAGATCGCCAATATCAA	58076 - 58095	
	CT050/51-7[f]	AGTCACTCCAGACAATTCTA	57699 - 57718	

	CT050/51-8[f]	TTAGTGAGACAGGCATTGA	57160 - 57178	
CT058/CT059	CT058/9-1[b]	AGTACCGGCCGAATCTCTTTCTCC	67315 - 67338	1634
(*fer*)	CT058/9-2[b]	GTCGGGGGTTCGAATCCCTCTA	68927 - 68948	
	CT058/9-3[f]	TCTCATTACTTCTCTTGCGT	68432 - 68451	
CT115-118[g]	---	---	---	---
CT119 (*incA*)	CT119-1[b]	GGGGCATTCGATGATATAAATAAG	136695 - 136718	1273
	CT119-2[b]	TGCACTGCGCAGACAAGAA	137949 - 137967	
CT143	CT143-1[b]	AAGGGCGGAGACTATGTGGACA	159770 - 159791	1325
	CT143-2[b]	AGCGCTGAACGCATACTTATTTTA	161071 - 161094	
CT144[g]	---	---	---	---
CT147[h]	CT147-1[b]	GGGGAAAGTGAGCTCTTCGGTATC	165430 - 165453	2819
	CT147-2[b]	CGCCGCTACAACAGCTTTAGTGA	168226 - 168248	
	CT147-3[b]	ATTGCGTCCCAAGATATACGACAG	167928 - 167951	2294
	CT147-4[b]	CACGCCAACCCAGAATCCTT	170202 - 170221	
	CT147-5[f]	TAATCATCCACTAGAAGCG	166229 - 166247	
	CT147-6[f]	TGTTCTAGCTGCTCTTGAAT	167649 - 167668	
	CT147-7[f]	TGGATGGTGTTGCAGAATTA	168511 - 168530	
	CT147-8[f]	TACCTCTAGATGTTTTGCGT	169786 - 169805	
CT192	CT192-1[b]	ATATGCGCAAGCACACCTTCC	215717 - 215737	1016
	CT192-2[b]	CTGGGCGTCCATTCACAACA	216713 - 216732	
	CT192-3[f]	CGTATCGATTCCTTCTTCTA	215998 - 216017	
	CT192-4[f]	CTCCTCTTATTGAAGAAGCT	2156196 - 216215	
CT195	CT195-1[b]	CCTCCGCCTAATCCTCGACTACAT	219655 - 216678	1320
	CT195-2[b]	CCAGCGGTTGATATTTCTTGATTA	220951 - 220974	
CT214	CT214-1[b]	AGGGCTTCTATTCCTCAAACAGTA	241463 - 241486	1797
	CT214-2[b]	TTCCCCGTTCTAAAGATCAGTTAT	243236 - 243259	
	CT214-3[f]	AACAGCCTGGATCTATATCA	242027 - 242046	
CT223	CT223-1[b]	GCAACGCATATCGCTCCTCA	251102 - 251121	1308
	CT223-2[b]	GTGCGCCCCTTCTCGTAAAG	252390 - 252409	
CT228/CT229	CT228/9-1[b]	CGGTCCCGGATTATCAAAACAAGT	254667 - 254690	1816
	CT228/9-2[b]	ATGCGGCCATCCCAGAAGC	256464 - 256482	
	CT228/9-3[f]	AGATTACGCAAACGTTGCTC	255033 - 255052	
	CT228/9-4[f]	GTTGTGATTGCAGCAGTAG	255972 - 255990	
CT232/CT233	CT232/3-1[b]	GATTAGGCGGAGGGGTTCTCTT	259223 - 259244	1262
(*incB*/*incC*)	CT232/3-2[b]	CTCTCCGCGACGCAAACTAAG	260464 - 260484	
CT249	CT249-1[b]	ACCACCCTTTAGCCATCCATTCC	279170 - 279192	704
	CT249-2[b]	AATTGCGCCGCCTCCTTGTA	279854 - 279873	
CT288	CT288-1[b]	TTTTACGCACAATGAACCCAGAAA	321582 - 321605	2063
	CT288-2[b]	CGGGCTCCTCGGGAACAG	323627 - 323644	
	CT288-3[f]	TTACCTGACCTCAGACACC	322164 - 322182	
	CT288-4[f]	GTCAGCTCGTCGTTTATTG	323121 - 323139	
CT293 (*accD*)	CT293-1[b]	TGCGCCAGAAGCTCCAGAAGTAGC	326322 - 326345	1350
	CT293-2[b]	AGGATCTGGCTGGGGATGGTTAGC	327648 - 327671	
CT365	CT365-1[b]	AAATTCGCAAACTTGCTCTTTTTC	416102 - 416125	2213
	CT365-2[b]	GATCGGGATTCCCCTGGATA	418295 - 418314	
	CT365-3[f]	CTAACTTCCAAGTTTCCTCT	416572 - 416591	
	CT365-4[f]	CTCATTGCAGGTATTGTTGT	417720 - 417739	
CT412 (*pmpA*)[i]	---	---	---	---
CT413 (*pmpB*)[i]	---	---	---	---
CT414 (*pmpC*)[j]	---	---	---	---
CT442 (*crpA*)	CT442-1[b]	CTCCTCCCTTTCCATACATCATCT	511256 - 511279	783
	CT442-2[b]	AAGCGATTCTTTCTTCCGATACAT	512015 - 512038	
CT443 (*omcB*)[k]	---	---	---	---
CT456 (*tarp*)	CT456-1[b]	ACAAACGTTACCCGGTATGCTGTT	530723 - 530746	3362
	CT456-2[b]	TTGCGCCTTGTCGATTGTGAT	534064 - 534084	
	CT456-3[f]	TACCTCATCAAGCGATCATA	531252 - 531271	
	CT456-4[f]	CCACCAGTTGTTATTATGTC	533470 - 533489	
	CT456-5[f]	AGACATGTCTCTTCCTTCAT	531867 - 531886	
	CT456-6[f]	TACATCAGAGATTACGTCTC	532889 - 532908	
	CT456-7[f]	GAGTTTCATTGGAGAAGGAA	532413 - 532432	
	CT456-8[f]	CGTTACCCGGTATGCTGTT	530728 - 530746	
	CT456-9[f]	TACAAACACTACTGCCTTCA	533363 - 533382	
	CT456-10[f]	TTGTTACTACCTACGTCATC	531328 - 531347	
	CT456-11[f]	CTAATTAAATCGGCTGTTG	530869 - 530887	
CT529	CT529-1[b]	ACGCGGCTCCTTAAAGCAAACAA	596464 - 596486	1659
	CT529-2[b]	CGCGCATATCCGGGGAGTCT	598103 - 598122	
	CT529-3[f]	TCTCGCAAGCATTTTCCTCT	596984 - 597003	
CT618	CT618-1[b]	TCCCGATATGCCTCCTTTTGAGTC	698080 - 698103	1360
	CT618-2[b]	ATGCGCACGCAAGCCAATC	699421 - 699439	
CT622	CT622-1[b]	GGCTCCCCCTCAATTCACAAACTT	707046 - 707069	2319
	CT622-2[b]	GGTCGCGGAAACCAAATGAAATA	709342 - 709364	
	CT622-3[f]	TGATTGCTTGTATTTCGGCT	707784 - 707803	

	CT622-4[f]	TTCAGCATCGTCCTCTGTAA	708885 - 708904	
	CT622-5[f]	AGAAGAGATTATGCAGAAGC	707298 - 707317	
CT623	CT623-1[b]	TTTGCCCATTAATAATTGGATTCA	708957 - 708980	1516
	CT623-2[b]	CATGGGTCGGTTGTATGAGATGT	710450 - 710472	
CT624 (*mviN*)	CT624-1[b]	ATCGAGCGCGGACATCTCC	710225 - 710243	2326
	CT624-2[b]	GATATCGCGCCCTTCGTAAAGA	712529 - 712550	
	CT624-3[f]	CTAAAGGTTTTGTTGGGACT	710827 - 710846	
	CT624-4[f]	ACGTGCGCAAAATATTTCTC	712033 - 712052	
CT653 (*yhbG*)	CT653-1[b]	AGCCGCGATAGCTAACGAAGTG	750194 - 750215	922
	CT653-2[b]	GAAGGCGGAATGAAAGTCCTCTC	751093 - 751115	
CT674 (*yscC*)	CT674-1[b]	TTTCAAGCGGAATCGCAAGGAAT	770234 - 770256	3229
	CT674-2[b]	CCGGGATCGAACCGACGAC	773444 - 773462	
	CT674-3[f]	AGAGCCATCAGATTTTCTCT	770846 - 770865	
	CT674-4[f]	AGAGGAAGAGAACTGAGTAA	772996 - 773015	
	CT674-5[f]	ATGACTTGAAAGTCGTTGAA	771470 - 771489	
	CT674-6[f]	TCCGTACATCATATCACTGA	772505 - 775524	
	CT674-7[f]	GATATCGGAGTCAATCTTGTT	772771 - 772791	
CT675 (*karG*)/	CT675/6-1[b]	CCCGGCTTTGGGCATTCC	773458 - 773475	1933
CT676	CT675/6-2[b]	TCATTCCGGTAACAGGGGTTCG	775369 - 775390	
	CT675/6-3[f]	CGTGATCAGATTAATCAGCT	774669 - 774688	
CT677/CT678/	CT677/9-1[b]	ATGGGGCCAGGACGGGTCTA	775420 - 775439	2393
CT679	CT677/9-2[b]	AAATTTTATCTCCGGTGCGTCCTG	777789 - 777812	
(*frr*, *pyrH*, *tsf*)	CT677/9-4[f]	GAAGATATCACTGTACCAAC	776088 - 776107	
	CT677/9-5[f]	TGGCTAAAGACATTGCTATG	777107 - 777126	
CT680 (*rs2* or *rpsB*)[c]	---	---	---	---
CT681 (*ompA*)[j]	---	---	---	---
CT682 (*pbpB*)[h]	pbpB-1 [b, c]	---	---	---
	pbpB-2 [b, c]	---	---	---
	CT682-1[b]	GCATTGTGATCGCGCAGGAGTA	780841 - 780862	3149
	CT682-2[b]	TTTCCGCCTCTTCCATAGTCGTTC	783966 - 783989	
	CT682-3[f]	TTGATAGGCAAGCGATCTAT	781484 - 781503	
	CT682-4[f]	ATATTCTCCAGGAAGTCCTA	783318 - 783337	
	CT682-5[f]	TGGAAATGTTTGAGTGTGAA	782079 - 782098	
CT683 (*TPR-motif protein*)	CT683-1[b]	TCGCTGCGGTAGGATATGAAGATG	783756 - 783779	1702
	CT683-2[b]	TCGCCGCGTAAATGAACCAAT	785437 - 785457	
	CT683-3[f]	GACTGTCAAACAGCCCTTAA	785108 - 785127	
CT684[h]	CT684-A1[b]	TGGGCATTACAATCTTGGGTTATG	784615 - 784638	1471
	CT684-A2[b]	AACAGCGGCATGCAGTTGATG	786065 - 786085	
	CT684-B1[b]	ATTTCGGGAGGCAATCCACAAT	785785 - 785806	1398
	CT684-B2[b]	TCCCGGGAATCCATATACCTCTTC	787159 - 787182	
	CT684-A3[f]	TCATCAATAGGGATGCCTAA	785651 - 785670	
	CT684-A4[f]	TCTGGATCTGCGTCTTCTAA	785609 - 785628	
CT685/CT686	CT685/6-1[b]	GCTCGGGAAGCAACCAAGTTATTA	786722 - 786745	2386
	CT685/6-2[b]	AAGCGAGTTCCCATGATACGAGAT	789084 - 789107	
	CT685/6-3[f]	TCTTGATGTCGATGCTTTGA	787308 - 787327	
	CT685/6-4[f]	ATGCAGAAGCTCGATTACTT	788502 - 788521	
CT687 (*yfhO_1*)[g]	---	---	---	---
CT688 (*parB*)[g]	---	---	---	---
CT689 (*dppF*)[g]	---	---	---	---
CT694	CT694-1[b]	TACAGGGGGAGGCGCTTCCTTA	796071 - 796092	1541
	CT694-2[b]	CGCGCTCTTCTAGCTCTCCCTCTT	796528 - 797551	
	CT694-3[f]	GATAACTCTTAACCCCATTG	796267 - 796286	
CT760 (*ftsW*)	CT760-1[b]	CTTGGGCCATTGCATTGAGTAAT	892695 - 892717	1417
	CT760-2[b]	CCCCAGAGAACATCCGATTGAC	894090 - 894111	
	CT760-3[f]	AGGATGTAGGTAAACTTGCA	893445 - 893464	
CT783	CT783-1[b]	AGCGGGGATTCAGCATTCCT	919328 - 919347	1427
	CT783-2[b]	TGCCCCTCGCCTCTTCATC	920676 - 920694	
	CT783-3[f]	TGTCAATACCTTCCCTAGTT	919913 - 919932	
CT812 (*pmpD*)[i]	---	---	---	---
CT813	CT813-1[b]	CTGCGTGTCGCTCTGGAAAATAAT	954988 - 955011	1460
	CT813-2[b]	AGGCCGAGCCCTACTCAAAAACT	956365 - 956387	
CT818 (*tyrP_2*)	CT818-1[b]	CCTGGCGGGAAAGGGACTCT	961600 - 961619	1531
	CT818-2[b]	GCGCATAATCGCGATCATACAATC	963107 - 963130	
	CT818-3[f]	GAATAGCACGTTTAACCTCA	962512 - 962531	
CT852 (*yhgN*)	CT852-1[b]	CTGCCGCACCAGCAAGGAT	1001213 - 1001231	768
	CT852-2[b]	TAGGCGCTCAACTTCTGGTATCTG	1001957 - 1001980	
CT859 (*ispH*)/	CT856/60-1[b]	GAGGGGGCTTTGCGGATTTAT	1011689 - 1011709	2779
CT860	CT856/60-2[b]	CCGGAATGCTTGGCTTGACA	1014448 - 1014467	
	CT856/60-3[f]	ACGACATTGAGTATGGATGA	1012253 - 1012272	
	CT856/60-4[f]	ACCCCGATATCTCATAAATC	1013829 - 1013848	

	CT856/60-5[f]	GTATTTCAGTTGCCTAAGGA	1012658 - 1012677	
CT861/CT862	CT861/2-1[d]	GAGGGCAGAGGCTTCTTCACAAG	1014122 - 1014144	2400
(*lcrH_2*)	CT861/2-2[b]	CTAGGCGTCCCAATTGGAGACTC	1016499 - 1016521	
	CT861/2-3[f]	AATAGCTCTCCAACCATCAA	1014640 - 1014659	
	CT861/2-4[f]	GACTATGAGGAAAGTTCTAC	1016071 - 1016090	
	CT861/2-5[f]	TCTCCTGTTGCTATTGTTTG	1014159 - 1014178	
	CT861/2-6[f]	TTGACTTTCTCCTGATGCTT	1015102 - 1015121	
CT867/CT868	CT867/8-1[d]	TCCCGACTGCTGGGGCTTAGA	1022902 - 1022922	2889
	CT867/8-2[d]	CATCGCGTCATGCCATGTCCTAT	1025768 - 1025790	
	CT867/8-3[f]	TCCTTCAGCTACTCTGATT	1022935 - 1022953	
	CT867/8-4[f]	TGGTAAGCGGATTACAGAT	1023641 - 1023659	
	CT867/8-5[f]	TTCATGGCGTTCTACAGAAT	1024302 - 1024321	
	CT867/8-6[f]	AATGTCAGAATCCCAAGCA	1024923 - 1024941	
	CT867/8-7[f]	AGAGGGAAGTCTTAATTTCC	1025542 - 1025561	
CT869 (*pmpE*)[i]	---	---	---	---
CT870 (*pmpF*)[i]	---	---	---	---
CT871 (*pmpG*)[i]	---	---	---	---
CT872 (*pmpH*)[i]	---	---	---	---
CT874 (*pmpI*)[i]	---	---	---	---

[a] Open reading frame (ORF) numbers, primer location (genome coordinates), primer sequences and amplicon size are based on the D/UW3 strain genome annotation (GenBank accession number NC_000117).
[b] Amplification primers also used for automated sequencing.
[c] Previously described in Gomes *et al.*, 2007, *Genome Res*, 17:50-60 (170).
[d] Primers exclusively used for PCR amplification.
[e] Primer sequences, location and amplicon size refers to the L2/434 strain genome annotation (GenBank accession number NC_010287) as these primers were designed for amplification and automated sequencing in LGV and ocular strains, and they have no homology in the D/UW3 genome sequence.
[f] Primers exclusively used for automated sequencing.
[g] Previously described in Nunes *et al.*, 2008, *Genome Biol*, 9:R153 (163).
[h] Due to the large gene size (for CT147 and CT682) or for PCR optimization (for CT684), two PCR primer pairs were designed to generate two overlapping amplicons for the entire gene.
[i] Previously described in Gomes *et al.*, 2006, *J Bacteriol*, 188:275-86 (111).
[j] Previously described in Gomes *et al.*, 2004. *J Bacteriol*, 186:4295-4306 (169).
[k] Previously described in Millman *et al.*, 2001. *J Bacteriol*, 183:5997-6008 (38).

Supplemental Table S3.1. Accession numbers of *C. trachomatis* genomic sequences used in this study.

Available at http://jb.asm.org/content/194/22/6143/suppl/DCSupplemental

Supplemental Table S3.2. Primers used in the analysis of transcriptional linkage (TL) and in the determination and analysis of the transcriptional start sites (TSSs).

Available at http://jb.asm.org/content/194/22/6143/suppl/DCSupplemental

Supplemental Table S3.3. Deletion and insertion events, and pseudogenes in *inc* genes of *C. trachomatis* [a].

***inc* gene**	**Deletions**	**Insertions**	**Pseudogenes**				
	position[a] (nt)	Nr. of codons	strains	position[a] (nt)	Nr. of codons	strains	strains
ct058	333	1	A/Har13	123	1	A/Har13	Ocular except A/Har13
	834	4	LGV except L1/440				
ct101							Urogenital except D/UW3, E/11023, and E/Bour
ct115[b]				before start codon	10	Ocular	
				before start codon	14	LGV	
				411	4	LGV	
				423	1	LGV	
ct135				197	1	E/SW3	A/2497, A/Har13, B/Jali29, F/SW4, and J/6276
ct147	279	1	A/363, A/2497, and A/5291				

ct192[c]	before start codon	26	LGV				B/Jali20[c]
	before start codon	67	Ocular				
	210	8	D/SotonD1 and all E and F strains				
ct214				1110	1	Ocular	
ct222	222	1	LGV				
ct223	399	1	LGV				
	786	1	LGV				
ct226[d]	21	2	ocular				
ct227							A/Har13
ct228							B/Jali20, B/TZ1A828/OT, D/SotonD6, and L1/115
ct288				1419	1	LGV	
ct300							LGV
ct358[e]	477	1	B/Jali20				
	492	3	LGV				
	519	5	LGV				
ct383	717	4	A/363 and A/7249				
ct442	162	1	LGV	66	2	All except D/SotonD5, D/SotonD6, and G, Ia, J, and K strains	
	213	1	LGV, E and F strains, and Soton D1				

[a] The position is based in the relative *inc* gene sequence of *C. trachomatis* D/UW3. Whenever different start codons for the same inc gene were annotated in different strains, we followed the annotations from archetype ocular (A/Har13), urogenital (D/UW3), and LGV strains (L2/434) (except *ct226* and *ct358*, as indicated).

[b] Considering that *ct115* (*incD*) from urogenital strains have a different start codon from ocular and LGV strains, as annotated for A/Har13, D/UW3, and L2/434/Bu; however, if the start codon of *ct115* is identical for all strains and corresponds to the one annotated for C. *trachomatis* D/UW3 (as suggested by an upstream strong putative ribosome binding site [GUGAGG-N3-AUG; predicted start codon underlined] that is present in all strains in the same position), then *ct115* is a pseudogene in ocular strains

[c]Considering that *ct192* from ocular, urogenital and LGV strains have a different start codon, as annotated for A/Har13, D/UW3, and L2/434/Bu; however, similarly to *ct115/incD*, if the start codon of *ct192* is identical for all strains and corresponds to the one annotated for C. *trachomatis* L2/434/Bu (as suggested by an upstream strong putative ribosome binding site [AGGAGG-N5-GUG; predicted start codon underlined] that is present in all strains in the same position), then *ct192* is a pseudogene in ocular strains. This is the predicted start codon of *ct192* annotated in A/2497, B/Jali20, and B/TZ1A428/OT (166), and *ct192* is indicated as a pseudogene in these ocular strains (166).

[d] Considering that the start codon of *ct226* is in all cases the one annotated for strain A/Har13. Otherwise, assuming the start codons annotated for strains D/UW3 and L2/434/Bu for urogenital and LGV strains, then ocular strains have a 5 codon- and LGV strains an 11 codon-deletion relative to the start codon of urogenital strains, and *ct226* is a pseudogene in strain D/SotonD6.

[e] *ct358* is annotated as a pseudogene in L2/434 and its nucleotide sequence is 100% identical among LGV strains. However it is likely that in LGV strains this gene encodes a protein that is 8 amino acid residues shorter at its C-terminus than CT358 in ocular and urogenital strains [as annotated for LGV strain L2c (175)].

Supplemental Table S3.4. Distribution of the segregation into disease groups displayed by amino acid-based phylograms of Inc proteins, Pmps, and housekeeping proteins (HKs).

Segregation	Incs	Pmps	HKs
Tropism	CT115, CT135, CT214, CT223, CT383	PmpF, PmpH	
Ocular & Urogenital	CT232		
LGV & Ocular	CT116, CT195, CT224,	PmpE, PmpI	YraL

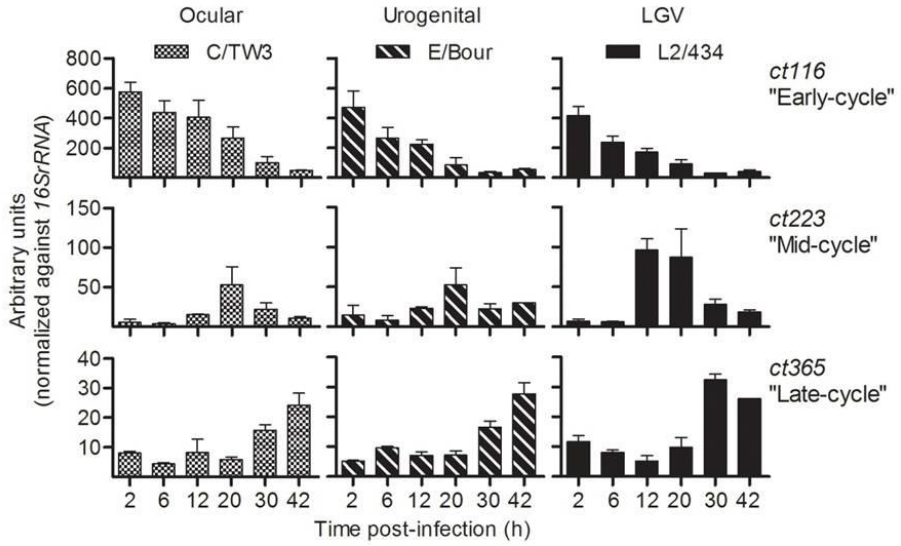
	CT442, CT850		
LGV & Urogenital	CT226		
Ocular	CT225		
LGV	CT006, CT036, CT101, CT117, CT118, CT146, CT179, CT192, CT196, CT222, CT227, CT228, CT229, CT249, CT288, CT345, CT357, CT358, CT365, CT449, CT484, CT565, CT618, CT813	PmpB, PmpC, PmpD, PmpG	Fer, AccD, PyrH, Rs2
None	CT005, CT058, CT119, CT134, CT164, CT300, CT440, CT483, CT728, CT789	PmpA	KarG, Frr, Tsf, YfO_1

Supplemental Table S3.5. Non-synonymous (d_N) and synonymous (d_S) substitutions between *inc*, *pmp*, and housekeeping (HK) genes among 51 *C. trachomatis* strains, and test of selection using the codon-based Z test (performed with MEGA5; www.megasoftware.net) as described in Materials and Methods. When $d_N/d_S > 1$, we tested for positive selection. If the results of the codon based Z-tests of selection were significant ($p < 0.05$), we further tested for lack of neutrality ($d_N \neq d_S$). Genes were considered likely under putative positive selection, only if both positive selection and neutrality tests were significant. The nucleotide sequences were obtained from the genomes of the 51 *C. trachomatis* strains listed in Supplemental Table S3.1.

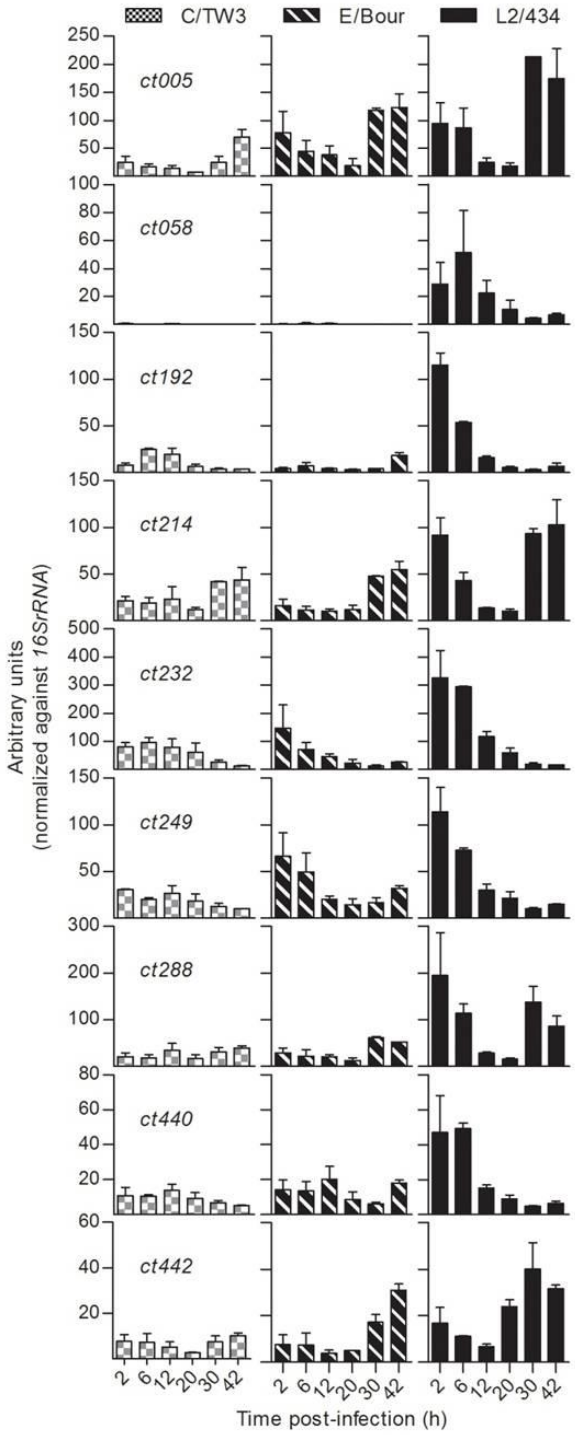
Excel file available at: http://jb.asm.org/content/194/23/6574/suppl/DCSupplemental

Supplemental Table S3.6. Real-time quantitative PCR (RT-qPCR) of mRNA levels of inc genes during the developmental cycle of *Chlamydia trachomatis*. The data shows for each inc at each time point the mean values (x105) ± SEM, normalized against the corresponding values of the 16SrRNA, from at least two independent experiments (except where indicated not determined, ND).

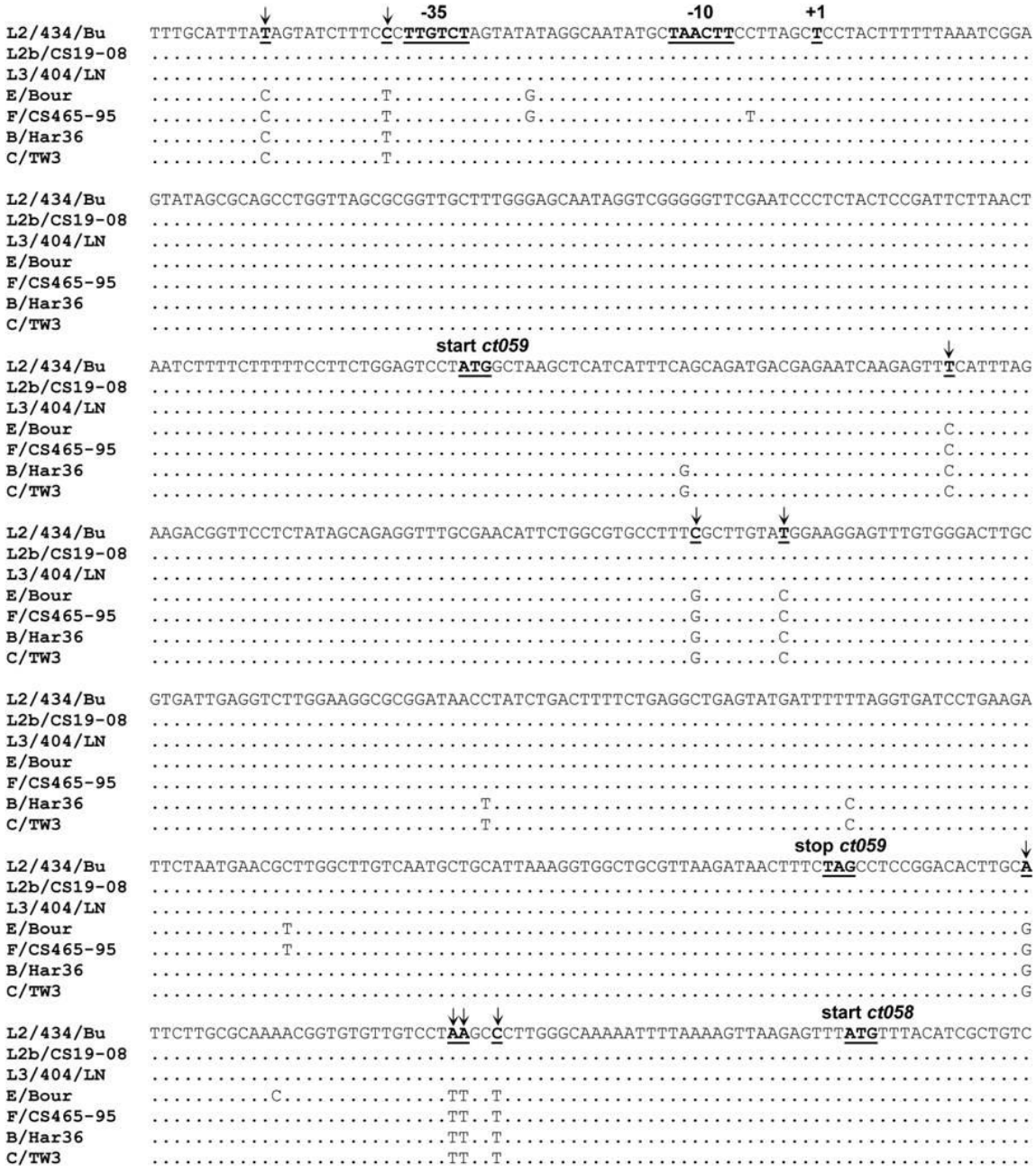
Excel file available at: http://jb.asm.org/content/194/23/6574/suppl/DCSupplemental



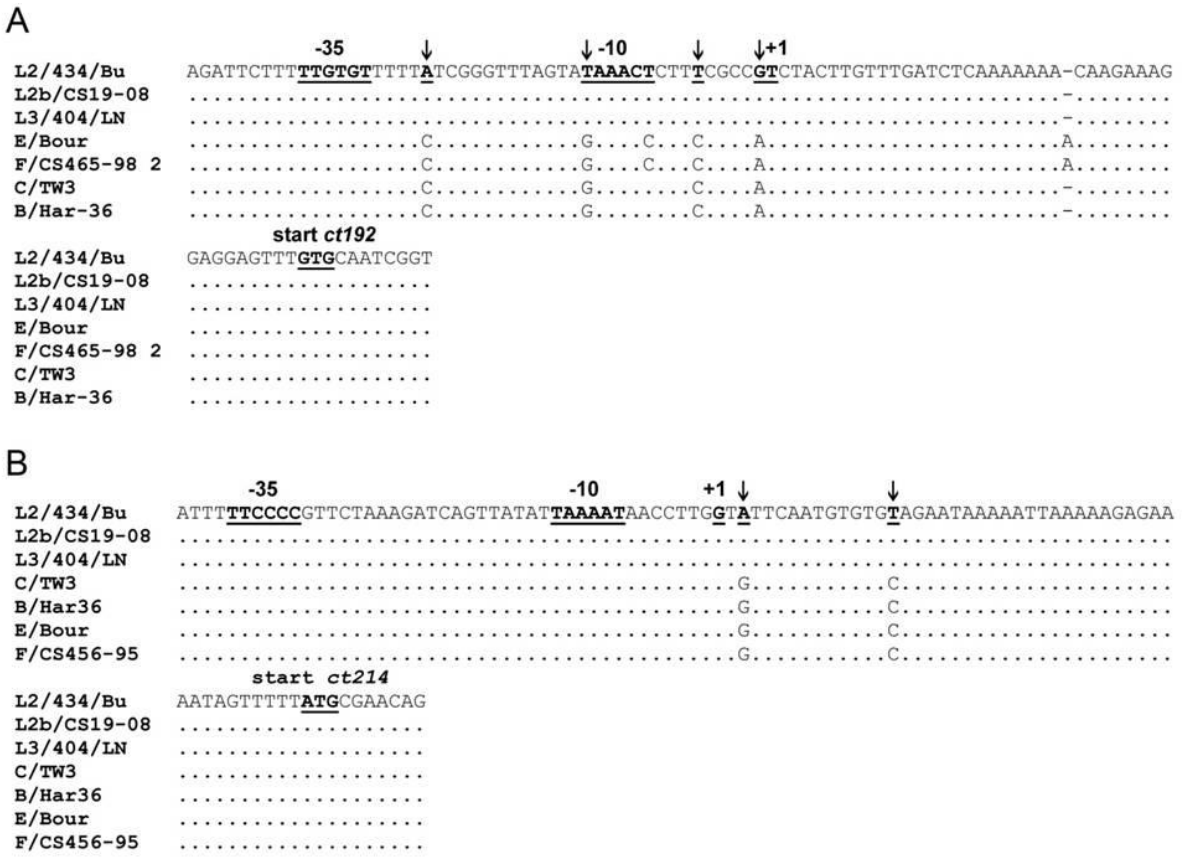
Supplemental Figure S3.1. Examples of variation of mRNA levels (profiles of expression) of *inc* genes throughout the developmental cycle of the indicated *C. trachomatis* strains. The expression values (mean ± SEM) result from the RT-qPCR raw data ($x10^5$) of each gene normalized against the *16SrRNA*, from at least two independent experiments.



Supplemental Figure S3.2. Differences in the mRNA levels of *inc* genes throughout the developmental cycle of *C. trachomatis* C/TW3, E/Bour, and L2/434. The expression values (mean ± SEM) of the indicated *inc* genes result from the RTqPCR raw data (x10^5) of each gene normalized against the *16SrRNA*, from at least two independent experiments.



Supplemental Figure S3.3. LGV-specific nucleotide differences in the promoter region of *ct059*, coding sequence of *ct059*, *ct059-ct058* intragenic region, and first codons of *ct058*. The nomenclature of D/UW3 is used. A dot "." indicates an identical nucleotide among all strains. The following features of strain L2/434 are depicted in the sequence (underlined and bold): the transcriptional start site of *ct059* determined by RACE (+1), the predicted start and stop codons of *ct059*, the deduced -10 and -35 hexamers within the promoter of *ct059*, the predicted start codon of *ct058*, and LGV specific nucleotide differences (arrows).



Supplemental Figure S3.4. LGV-specific nucleotide differences in the promoter region of *ct192* (A) and of *ct214* (B). The nomenclature of D/UW3 is used. A dot "." indicates an identical nucleotide among all strains. The following features of strains L2/434 are depicted in the sequence (underlined and bold): the transcriptional start sites of *ct192* and *ct214* determined by RACE (+1), the predicted start codons of *ct192* and *ct214*, the deduced -10 and -35 hexamers within the promoters of *ct192* and *ct214*, and LGV specific nucleotide differences (arrows).

Supplemental Table S4.1. Plasmids used and constructed in this work.

Available at: http://www.biomedcentral.com/1471-2180/14/40/additional

Supplemental Table S4.2. Primers used in this work for construction of plasmids.

Available at: http://www.biomedcentral.com/1471-2180/14/40/additional

Supplemental Table S4.3. Summary of results obtained in analyses of T3S signals in proteins of *Chlamydia trachomatis* and comparison to *in silico* prediction methods.

Available at: http://www.biomedcentral.com/1471-2180/14/40/additional

Supplemental Table S6.1. List of *Chlamydia trachomatis* LGV strains used for comparative genomic analyses.

Strain name	Source	Year	Country	GenBank Accession numbers (chromosome/plasmid)	Reference
L1/1322/p2	Genital ulcer	1995	South Africa	HE601951 / HE603219	(37)
L1/115	LGV patient	<1994	South Africa	HE601952 / HE603218	(37)
L1/224	LGV patient	<1994	South Africa	HE601953 / HE603220	(37)
L1/440/LN	Lymph	1968	USA	HE601950 / X06707	(37)
L2/25667R	Rectal biopsy	<1983	USA	HE601954	(37)
L2/434/Bu	Lymph node	1968	USA	NC_010287 / AM886278	(46)
L2b/UCH-1	Rectum	2006	UK	NC_010280.2 / AM886279	(46)
L2b/8200/07	Rectum	2007	Sweden	HE601795 / HE603222	(37)
L2b/UCH-2	Rectum		UK	HE601956 / HE603227	(37)
L2b/Canada1	Rectum	2004	Canada	HE601963 / HE603223	(37)
L2b/Canada2	Rectum	2005	Canada	HE601957 / HE603224	(37)
L2b/LST	Rectum	2008	France	HE601958 / HE603226	(37)
L2b/CV204	Rectum	2006	France	HE601960 / HE603225	(37)
L2b/795	Rectum	2004	France	HE601949 / HE603221	(37)
L2b/Ams1	Penile ulcer	2004	Netherlands	HE601959 / HE603213	(37)
L2b/Ams2	Anus	2005	Netherlands	HE601961 / HE603214	(37)
L2b/Ams3	Anus	2004	Netherlands	HE601962.1 / HE603215	(37)
L2b/Ams4	Anus	2005	Netherlands	HE601964.1 / HE603216	(37)
L2b/Ams5	Anus	2004	Netherlands	HE601965.1 / HE603217	(37)
L2b/CS19/08	Anus/Rectum	2008	Portugal	CP009923 / CP009924	This work
L2b/CS784/08	Anus/Rectum	2008	Portugal	CP009925 / CP009926	This work
L3/404/LN	Lymph node	1967	USA	HE601955 / HE603228	(37)

Supplemental Table S6.2. Genetic variability among LGV strains.

Excel file available at: http://www.sciencedirect.com/science/article/pii/S1567134815000696

Supplemental Table S7.1. Summary of sequencing and assembly data.

Avaliable at http://www.sciencedirect.com/science/article/pii/S1567134813001081

Supplemental Table S7.2. Variable nucleotide positions among clones from the E/CS88(i) population.

Polymorphism location [a]	*loci* [b]	Variable nucleotide (low ↔ **high frequent**) [c]	aa change
123302	CT106 (*yceC*)	C ↔ **T**	silent
152891	CT135	G ↔ **A**	Trp ↔ Stop codon
202869	CT180 (*tauB*)	T ↔ **G**	Phe ↔ Leu
222530	CT197 (*gcp_1*)	C ↔ **T**	silent
231331	CT205 (*pfkA_1*)	T ↔ **G**	silent
298698	CT266	A ↔ **G**	silent
325352	CT289	C ↔ **T**	Ser ↔ Leu
382012	IGR CT333-CT334	G ↔ **A**	---
409481	CT356 (*yyaL*)	A ↔ **G**	silent
451617	CT394 (*hrcA*)	G ↔ **A**	silent
507097	CT437 (*fusA*)	T ↔ **C**	Phe ↔ Ser
555161	CT479 (*oppB_2*)	C ↔ **T**	Thr ↔ Ile
588962	CT511 (*rlpO*)	A ↔ **G**	Tyr ↔ Cys
598692	CT530 (*fmt*)	G ↔ **A**	Glu ↔ Lys
599233	CT530 (*fmt*)	C ↔ **G**	silent
652385	CT579	A ↔ **G**	Ile ↔ Val
749629	CT652 (*recD_2*)	G ↔ **A**	silent
763654	CT664	T ↔ **C**	Tyr ↔ Hys
819550	CT709 (*mreB*)	C ↔ **T**	Ala ↔ Val
928175	CT792 (*mutS*)	C ↔ **T**	silent
975533	CT827 (*nrdA*)	C ↔ **T**	silent
979576	CT829 (*trmB*)	T ↔ **C**	silent
997865	CT847	T ↔ **C**	silent

[a] Polymorphism location refers to the location in the reference E/Bour chromosome (GenBank accession number HE601870.1).

[b] The *loci* designations are based on genome annotation of the strain D/UW3 (GenBank accession number NC_000117). IGR, intergenic region with adjacent ORFs indicated.

[c] The nucleotide changes in open reading frames are presented in the 5' to 3' direction. Boldface type refers to the high-frequent nucleotide in the clone-mixture population, which is supported by more than 70% of the respective reads.

Supplemental Table S8.1. Sequence data from the strains analyzed in this study.

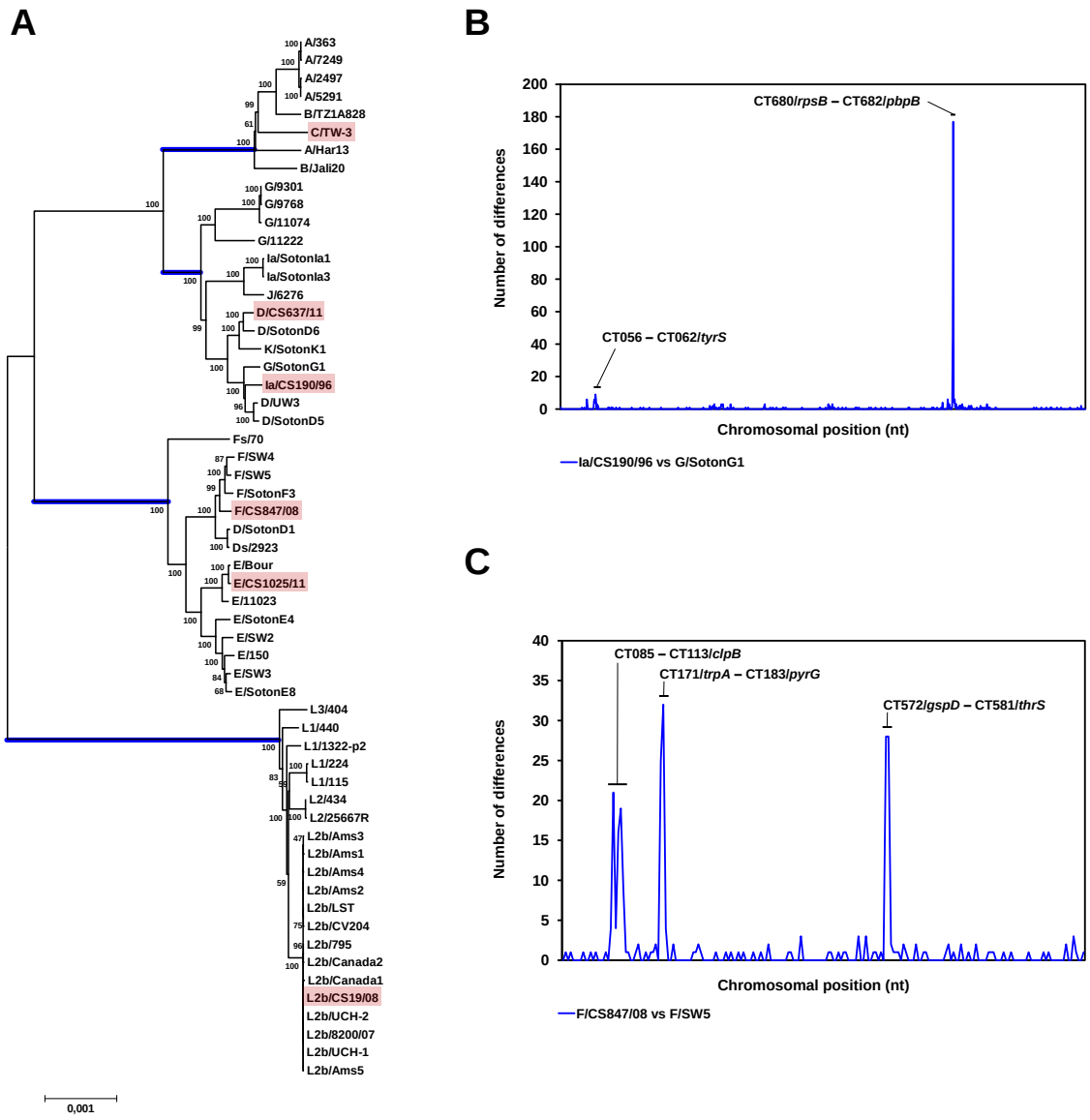
Strain	Chromosome (bp) [a]	Plasmid (bp) [a]	**Mean Depth Coverage Plasmid/Chromosome (ratio)**					**GenBank accession number [a]**	
			P5-7	P20	P30	P50	P100	Chromosome	Plasmid
C/TW-3	1043554	7501	6717/1117 (6,01)	---	---	---	1823/262 (6,96)	CP006945	CP006946
D/CS637/11	1042680	7493	824/179 (4,60)	642/74 (8,68)	2318/266 (8,71)	---	---	CP007131	CP007132
E/CS1025/11	1043034	7502	6513/464 (14,03)	1520/104 (14,62)	2935/319 (9,20)	---	---	CP010567	CP010568
F/CS847/08	1043060	7493	3116/444 (7,02)	195/41 (4,76)	5727/1137 (5,04)	---	---	CP010569	CP010570
Ia/CS190/96	1042034	7471	4085/307 (13,31)	373/56 (6,66)	1164/139 (8,37)	3465/551 (6,29)	2107/247 (8,53)	CP010571	CP010572
L2b/CS19/08	1038864	7500	586/68 (8,62)	875/139 (6,29)	1258/104 (12,10)	---	---	CP009923	CP009924

[a] Both the sequence sizes and GenBank accession numbers refer to the chromosome and plasmid sequences of the major abundant clone in the first sequenced population.

Supplemental Table S8.2. Oligonucleotide primers used in PCR and qPCR assays.

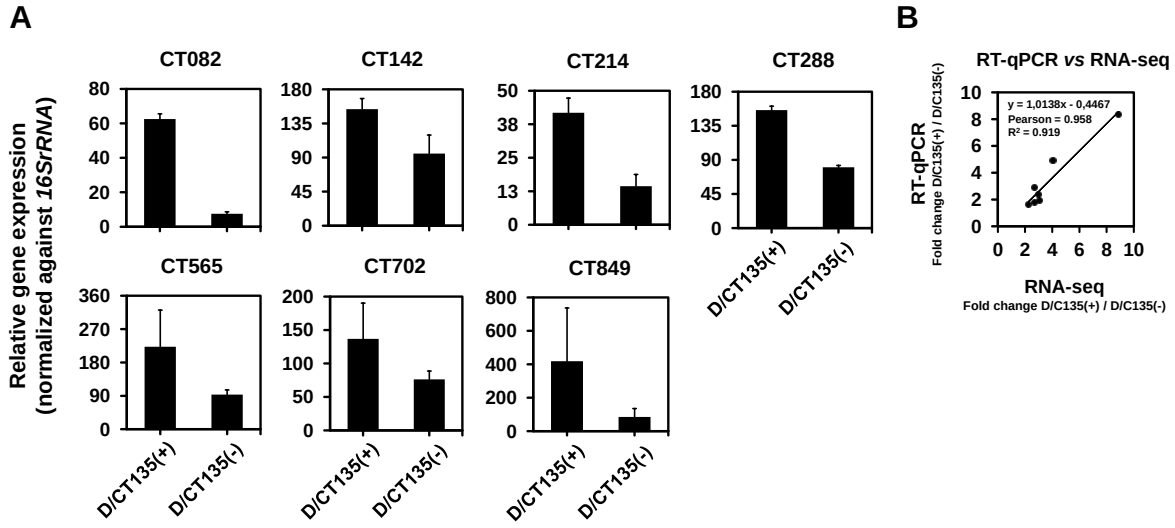
ORF [a]	Primers	Primer sequence (5' to 3')	Primer location	Amplicon (bp)
RNA-seq vs RT-qPCR assays				
CT082	CT082-A	CAGCTCATGCAGCCAAAGAG	94352-94371	52
	CT082-B	GACGTTTCCGTAGCCTTGGAT	94383-94403	
CT142	CT142-A	ATGCTTGCGAAAGGTGTGTACA	159509-159530	51
	CT142-B	CCATCGCTTCCCAACTCCTA	159540-159559	
CT214	CT214-A	TTATTTCCGGACAAGCAGATGA	242770-242791	51
	CT214-B	TTTTGATCCCAATCCGATTAGG	242741-242762	
CT288	CT288-A	GCCTGCCTTTTATCGCTGTTAT	322619-322640	51
	CT288-B	CCATCCCCAATGCTAAGGAA	322650-322669	
CT565	CT565-A	CTATTACGCTAAAAACGGGCCTAT	636539-636562	51
	CT565-B	GCAACCCCGCATACCAAAG	636512-636530	
CT702	CT702-A	TTCTTCAGCCAGATTTGCTACTCA	807888-807911	52
	CT702-B	GCATGAATCAGCTTTTCCACATTA	807860-807833	
CT849	CT849-A	CAACAGCAATTAAACCAAGAAACG	998457-998480	51
	CT849-B	GCCAACGACAGCGTATTTGATT	998430-998451	
CTr01-04	*16SrRNA*-9 [c]	GCGAAGGCGCTTTTCTAATTTAT	854857-854879	76
	16*SrRNA*-10 [c]	CCAGGGTATCTAATCCTGTTTGCT	854909-854932	
mRNA decay assay (qPCR)				
CT134	CT134-A	CGGTTACGATGAGATTTGTTGTAGA	151694-151718	52
	CT134-B	CGCCACTACTTTTTCTGCAGTCT	151723-151745	
CT135	CT135-A	ACGAACGGATCATGTTTGAAGA	152387-152408	51
	CT135-B	CGGCTTCGAGAACACTAGGAACT	152415-152437	
Growth rate/doubling time and mRNA decay assays (qPCR)				
CT681/*ompA*	OmpA-9 [c]	TGCCGCTTTGAGTTCTGCTT	780009-780028	76
	OmpA-10 [c]	GTCGATCATAAGGCTTGGTTCAG	779953-779975	

[a] Open reading frame (ORF) numbers are based on the D/UW3 strain genome annotation (GenBank No. NC_000117).
[b] Based on the chromosome sequence of D/UW3 strain (GenBank No. NC_000117).
[c] Previously described in J.P.Gomes et al (Microbes Infect. **7:** 410-420, 2005) (251).



Supplemental Figure S8.1. Genome make-up of the studied *C. trachomatis* strains and evaluation of putative mosaic structures. Panel A. Evaluation of the genetic backbone of the studied strains in the frame of the species phylogeny and diversity. The phylogenetic tree was constructed [using MEGA5 (233)] based on a core-genome alignment enrolling 52 previously analyzed strains (37) and the six strains evaluated in the present study (highlighted in red) by using the neighbor-joining method with bootstrapping (1000 replicates) (234) based on distance estimates using a Kimura two-parameter model for substitution events (235). This analysis shows that the strains evaluated in the present experimental evolution study represent not only the three disease groups (ocular, epithelial-genital and LGV) but also the four major genetic branches of the species tree in *C. trachomatis* (highlighted in blue): ocular clade (serovars A-C), prevalent epithelial-genital clade (mainly serovars E and F), non-prevalent epithelial-genital clade (serovars D, G, Ia, J, and K) and LGV clade (serovars L1-L3, and L2b). Panel B. Recombination analysis of the Ia/CS190/96 strain. The graph shows the SNP density across the chromosome when comparing the sequences from Ia/CS190/96 with G/SotonG1 (a probable parental strain). Polymorphism was assessed through DnaSP v5 analysis (364) using a window size and a step size of 1000 base pairs each. The Ia/CS190/96 is a recombinant strain likely resulting from the genetic import of the typing gene CT681/*ompA* and surrounding region (estimated crossovers occurring in the gene CT680/*rpsB* and in the IGR

CT681/*ompA*-CT682/*pbpB*) from a serovar Ia parental strain to a strain with a similar genetic backbone of the serovar D or G strains clustering in the same sub-branch. The genetic backbone of the parental strain is likely more similar to that of G/SotonG1, as these two strains also share an unusual genetic feature at the plasmid (three and not four tandem 22-bp repeats at the ORI). The serovar Ia strain seems to have also imported a ~6000 bp region (enrolling the genes CT056 to CT062/*tyrS*) from a serovar E-like parental strain. Panel C. Recombination analysis of the F/CS847/08 strain. The graph shows the SNP density across the chromosome when comparing the sequences from the F/CS847/08 and F/SW5 strains. Polymorphism was assessed through DnaSP v5 analysis using a window size and a step size of 5000 base pairs each. The F/CS847/08 chromosome displays three putative recombining fragments enrolling the genomic regions from the genes: *i*) CT085 to CT113/*clpB* (with the deduced crossovers falling within the gene CT085 and a region spanning from the IGR CT107/*mutY* - CT108/*ybgI* to the CT113/*clpB* gene); *ii*) CT171/*trpA* to CT183/*pyrG* (putative crossovers occurring within the IGR CT171/*trpA*-CT172 and the gene CT183/*pyrG*); and *iii*) CT572/*gspD* to CT581/*thrS* (deduced crossovers likely occurred within a region enrolling half of the CT572/*gspD* gene and the 5'-end of the gene CT573, and within the CT581/*thrS* gene). The first two exchanged fragments were putatively inherited from a parental strain with a genetic backbone similar to the one from the strain E/Bour, whereas the third region matched the genome from a serovar G-like strain (closely related to the strains G/9301, G/9768 or G/11074).



Supplemental Figure S8.2. Confirmation of the RNA-seq results with RT-qPCR. Panel A. RT-qPCR of seven differentially expressed genes that were chosen in order to represent a wide range of values regarding both expression levels and expression fold-change in the D/CS637/11 CT135-positive *versus* D/CS637/11 CT135-negative comparison. The expression values (mean ±SEM) resulted from raw RT-qPCR data ($x10^5$) of each gene normalized to that of the *16S rRNA.* Panel B. Correlation between the fold-change expression determined by RT-qPCR *versus* RNA-seq. The Pearson correlation coefficient (0.958), the curve slope (1.014) and the best fit linear regression analysis (R^2 = 0.919) demonstrate the high degree of correlation between the RT-qPCR and RNA-seq analyses.