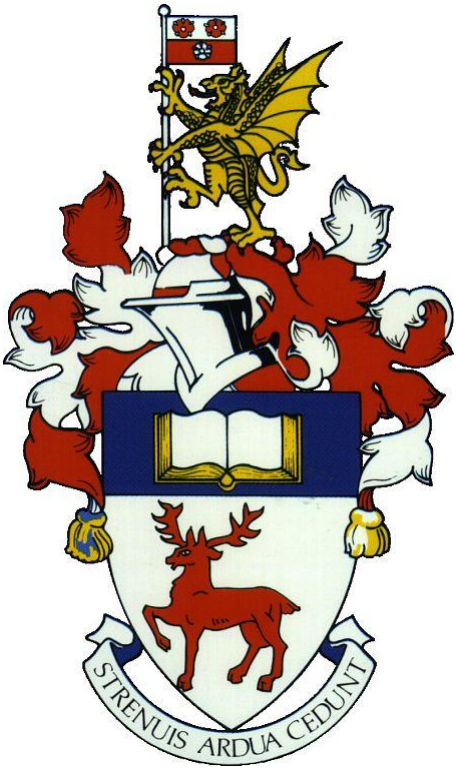
UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Clinical and Experimental Sciences



CHARACTERISATION OF GENETIC DIVERSITY WITHIN *CHLAMYDIA TRACHOMATIS* IN HIGH-RISK SEXUAL NETWORKS

By

Chloe Elizabeth Manning

ORCID 0000-0003-2428-7731

Thesis for the degree of Doctor of Philosophy

January 2023

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Thesis for the degree of Doctor of Philosophy

CHARACTERISATION OF GENETIC DIVERSITY WITHIN *CHLAMYDIA TRACHOMATIS* IN HIGH-RISK SEXUAL NETWORKS

Chloe Elizabeth Manning

Chlamydia trachomatis is the most commonly diagnosed bacterial STI worldwide. Transmission of *C. trachomatis* within sexual networks is facilitated by high-risk sexual behaviours. *C. trachomatis* infections are frequently asymptomatic which allows for un-interrupted *C. trachomatis* transmission within sexual networks. *C. trachomatis* strains are characterised by their *ompA* genotype: strains with *ompA* genotypes A – C are associated with trachoma, D – K with sexually transmitted urogenital infections, and L1 – L3 with Lymphogranuloma venereum (LGV). The *ompA* gene is a mutational hotspot within the *C. trachomatis* genome, and extensive sequence variation within *ompA* has been reported amongst *C. trachomatis* strains. *OmpA* genotyping was applied to *C. trachomatis* DNA extracted from a conjunctival swab collected from a 10 year old female that had migrated to the UK from Afghanistan, to ascertain whether the child was infected with a trachoma-causing strain, or a strain associated with sexually transmitted urogenital tract infections (Chapter 3). Sequencing of the *trpA* gene was also conducted, as polymorphisms within this gene have been shown to distinguish ocular and genital *C. trachomatis* strains (Chapter 3).

In recent years, multi-locus genotyping systems including multi-locus sequence typing (MLST), multi-sequence typing (MST), and multi-locus variable number tandem repeat (VNTR) analysis (MLVA), have been developed for *C. trachomatis*. Genotyping systems can elucidate diversity within *C. trachomatis* strains circulating in sexual networks. Multi-locus genotyping systems were evaluated to determine their suitability for application in a genotyping survey of LGV *C. trachomatis* strains in London (Chapter 4). The MLVA-*ompA* genotyping system was applied to genotype LGV-positive clinical specimens from London, to assess genotypic diversity within LGV *C. trachomatis* strains within this population (Chapter 5). The study detected several distinct LGV MLVA-*ompA* genotypes, in addition to a hybrid L2b/D *ompA* genotype that had previously been reported in Portugal. The data collected has added to the knowledge relating to the current LGV epidemic within MSM in the UK.

Multi-locus typing systems cannot achieve the resolution provided by whole genome sequencing. Whole genome sequencing can provide insight into the genomic diversity within LGV *C. trachomatis* strains. Three LGV isolates collected from asymptomatic and symptomatic MSM in Brighton were whole genome sequenced in Chapter 6, to identify any genomic loci or key sequences that may account for the differences in symptomology between the three patients. The study identified several mutations within the genomes of the three isolates. Of these, SNPs were detected within genes associated with intracellular replication of *C. trachomatis* in the two isolates collected from symptomatic MSM, that were absent from the isolate obtained from an asymptomatic MSM.

CONTENTS

TABLES

FIGURES

LIST OF ACCOMPANYING MATERIALS

Data DOI: https://doi.org/10.5258/SOTON/D2505

PUBLICATIONS AND PRESENTATIONS OF WORK PRESENTED IN THIS THESIS

Peer-Reviewed Publications

Mitchell, A., Patel, M., **Manning, C.**, Abbott, J. (2021). Reducing suspicion of sexual abuse in paediatric chlamydial conjunctivitis using *ompA* genotyping. *BMJ Case Reports*. 14(3):e238871.

Manning C., O'Neill, C., Clarke, I., Cliff, P., Rebec, M., Marsh, P. (2021). High-resolution genotyping of Lymphogranuloma Venereum (LGV) strains of *Chlamydia trachomatis* in London using multi-locus VNTR analysis-*ompA* genotyping (MLVA-*ompA*). *PLOS ONE.* 16(7):e0254233.

Poster Presentations

"High resolution genotyping methods rule out urogenital Chlamydia trachomatis infection in an eight-year-old child recently resident in Afghanistan." - **International Union against Sexually Transmitted Infections (IUSTI) World Congress**, Dublin, June 2018

"Genotyping methods used to identify infection in child from Afghanistan." - **Global Network for Anti-Microbial Resistance and Infection Prevention (NAMRIP) Festival of Early Career Research**, Southampton, June 2019

"*High resolution genotyping survey of Lymphogranuloma Venereum (LGV) strains of Chlamydia trachomatis in London using multi-locus VNTR analysis-ompA genotyping (MLVA-ompA)."* - **Faculty of Medicine Conference,** Southampton, June 2020

RESEARCH THESIS: DECLARATION OF AUTHORSHIP

I, Chloe Elizabeth Manning, declare that this thesis, entitled 'Characterisation of Genetic Diversity within *Chlamydia trachomatis* in High-Risk Sexual Networks', and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:-

Mitchell, A., Patel, M., **Manning, C.**, Abbott, J. (2021). Reducing suspicion of sexual abuse in paediatric chlamydial conjunctivitis using *ompA* genotyping. *BMJ Case Reports*. 14(3):e238871.

Manning C., O'Neill, C., Clarke, I., Cliff, P., Rebec, M., Marsh, P. (2021). High-resolution genotyping of Lymphogranuloma Venereum (LGV) strains of *Chlamydia trachomatis* in London using multi-locus VNTR analysis-*ompA* genotyping (MLVA-*ompA*). *PLOS ONE.* 16(7):e0254233.

Signature:

Date: 16th January 2023

ACKNOWLEDGEMENTS

I wish to firstly thank my supervisors, Professor Ian Clarke, Dr Peter Marsh, and Dr Colette O'Neill. Thank you for always being so generous with your time, for your continuous encouragement, and for providing me with this wonderful opportunity. I would also like to thank Dr Sarah Pearson and Rachel Skilton for their patience and kindness, and for their willingness to teach new lab techniques. I already miss our impromptu karaoke sessions and lab chats.

I would like to gratefully acknowledge Public Health England for the PhD studentship, and the Faculty of Medicine at the University of Southampton for the Vice-Chancellor's Scholarship; both provided the funding that made this PhD possible.

I would like to thank Dr Penelope Cliff and Monica Rebec for processing and anonymising the clinical specimens for the London LGV genotyping study. Thank you both for always being so helpful and willing to answer any questions. For providing guidance on various aspects of the LGV genotyping study; thank you to Dr Gary Whitlock at 56 Dean Street, Dr Achyuta Nori at Burrell Street Sexual Health Centre, and Dr John Saunders at Public Health England. Thank you also to our collaborators on the trachoma case study at Birmingham Women's and Children's Hospital: Dr Mitul Patel, Dr Joseph Abbott, Dr Alexander Mitchell, and Dr Helen Morris.

I would also like to thank every member of the vibrant, welcoming, and nurturing community that is the Molecular Microbiology department at the University of Southampton. I will miss our lunchtime conversations, where we discussed everything under the sun without any hesitation. A special thanks to Emma M, Emma C, and Muktar – I can't begin to express how thankful I am that we ended up in the same office these past few years. To say that this experience would not have been the same without you all is an understatement. Thank you for your endless wisdom, your encouragement, and for your friendship.

Finally, I would like to thank the family and friends who have been there to support me throughout this journey. There is no way that I could have got here without you. The past few years have been the most challenging years of my life, and your unwavering support and friendship has kept me going. Thank you for piecing me together again at my lowest points. To my parents – you deserve way more than a few lines here. You have been there cheering me on every step of the way. Your steadfast strength and your faith in me drives me to be and to do better. This is for you.

ABBREVIATIONS

×g	Times gravity
°C	Degrees Celsius
µg	Microgram
µL	Microlitre
µm	Micrometre
µM	Micromolar
3’	3 prime
4SP	4X sucrose phosphate
5’	5 prime
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BME	Black Minority Ethnic groups
bp(s)	Base pair(s)
BSA	Bovine Serum Albumin
CAP	Catabolite Activator Protein
CDS	Coding Sequence
COMC	Chlamydial Outer Membrane Complex
CTAD	Chlamydia Testing Activity Dataset
dH_2O	Distilled water
DMEM	Dulbecco’s Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dsDNA	Double-stranded DNA
EB	Elementary Body
ECACC	European Collection of Authenticated Cell Cultures
EDTA	Ethylenediaminetetraacetic acid
EQA	External Quality Assurance
ERGO	Ethics and Research Governance Online
ESSTI	European Surveillance of Sexually Transmitted Infections

FCS	Foetal Calf Serum
g	Gram
G6P	Glucose 6-phosphate
gDNA	Genomic DNA
GUM	Genitourinary Medicine
GUMCAD	Genitourinary Medicine Clinic Activity Dataset
HCV	Hepatitis C virus
HIV	Human Immunodeficiency Virus
HRA	Health Research Authority
HSPG	Heparan sulphate proteoglycan
IBD	Inflammatory Bowel Disease
IFN(-γ)	Interferon (-gamma)
IPTG	Isopropyl beta-D-1-thiogalactopyranoside
IRAS	Integrated Research Application Service
Kb	Kilo-base pairs
LB	Luria-Bertani
LGV	Lymphogranuloma venereum
M	Molar
Mb	Mega-base pairs
MCS	Multiple Cloning Site
MDA	Mass drug administration
mg	Milligram
MIF	Microimmunofluorescence
ml	Millilitre
MLST	Multi-Locus Sequence Typing
MLVA	Multi-Locus Variable Number Tandem Repeat Analysis
MOMP	Major Outer Membrane Protein
MSM	Men who have sex with men
MST	Multi-Sequence Typing
MTA	Material Transfer Agreement
MTOC	Microtubule Organising Centre

NAAT	Nucleic Acid Amplification Test
NCBI	National Centre for Biotechnology Information
ng	Nanogram
NHS	National Health Service
nm	Nanometre
nt	Nucleotide
NTC	No Template Control
NWLP	North West London Pathology
OD	Optical Density
OmcA	Outer membrane complex A protein
OmcB	Outer membrane complex B protein
OmpA	Outer membrane protein A
PBS	Phosphate buffer saline
PCR	Polymerase Chain Reaction
pg	Picogram
PHE	Public Health England
PID	Pelvic Inflammatory Disease
PmpH	Polymorphic membrane protein H
PrEP	Pre-exposure prophylaxis
RB	Reticulate Body
REC	Research Ethics Committee
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SNP	Single Nucleotide Polymorphism
ST(s)	Sequence Type(s)
STBRL	Sexually Transmitted Bacteria Reference Laboratory
STI	Sexually Transmitted Infection
T3SS	Type III Secretion System

TAE	Tris-Acetate EDTA
TarP	Translocated actin-recruiting phosphoprotein
TE	Trypsin-EDTA buffer
Tm	Melting Temperature
Tris	Tris(hydroxymethyl)aminomethane
TrpA	Tryptophan synthase alpha subunit
UHQ	Ultra High Quality
UV	Ultraviolet
VS	Variable Segment
VD	Variable Domain
VNTR	Variable Number Tandem Repeat
WGS	Whole Genome Sequencing
WHO	World Health Organisation
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

CHEMICAL FORMULAE USED IN THIS THESIS

$C_6H_8O_7$	Citric acid
$CaCl_2$	Calcium chloride
CO_2	Carbon dioxide
H_2O	Water
K_2HPO_4	Dipotassium hydrogen phosphate
$MgSO_4.7H_2O$	Magnesium sulphate heptahydrate
$NaNH_4HPO_4.4H_2O$	Sodium ammonium hydrogen phosphate tetrahydrate

CHAPTER 1 Introduction

1.1 History of Chlamydial Research

Chlamydiae are obligate intracellular Gram-negative bacteria that were first observed in 1907 by Halberstädter and Prowazek (Halberstadter and Von Prowazek, 1907). Halberstädter and Prowazek inoculated the conjunctivae of orangutans with conjunctival scrapings from human cases of trachoma ("rough eye"). Upon microscopic examination, particles were identified within intra-cytoplasmic inclusions near the nuclei of conjunctival epithelial cells. These inclusions were thought to represent colonies of the causative micro-organism, which was designated *Chlamydozoa*, from the Greek word *chlamys*, or mantle (Halberstadter and Von Prowazek, 1907). Similar inclusions were later identified in scrapings from adults with urethritis (Halberstadter and Von Prowazek, 1910; Lindner, 1910) and from infants with neonatal conjunctivitis (Lindner, 1909), implicating the same causative micro-organism. It was later thought that these inclusions were caused by a virus, as they could be passed through bacterial filters and early efforts to culture the agent had failed (Miyagawa *et al.*, 1935). In 1966, due to advances in electron microscopy and tissue culture techniques, *Chlamydiae* were re-classified as bacteria when it was found that they contained DNA and RNA, ribosomes, and cell walls (Moulder, 1966).

1.2 Taxonomy

The order Chlamydiales was established in 1971, and initially consisted of the family *Chlamydiaceae*, the genus *Chlamydia*, and the two species *Chlamydia psittaci* and *Chlamydia trachomatis* (Storz and Page, 1971). The two species were distinguishable on the basis that *C. psittaci* isolates (unlike *C. trachomatis*) were sulfadiazine-resistant and could not accumulate glycogen in their inclusions. A third species, *C. pneumoniae*, was added to the genus in 1989 (Grayston *et al.*, 1989), followed by *C. pecorum* in 1992 (Fukushi and Hirai, 1992). A further eight species have been distinguished in the years following. These include: *C. abortus*, which causes abortion in sheep, cattle, pigs, and occasionally humans (Everett, Bush and Andersen, 1999; Seth-Smith *et al.*, 2017); *C. avium*, a cause of respiratory disease in psittacines and pigeons (Sachse *et al.*, 2014); *C. caviae*, which causes conjunctivitis in guinea pigs (Read *et al.*, 2003); *C. felis*, a cause of conjunctivitis in cats (Sykes, 2005); *C. gallinacea*, which primarily infects poultry (Sachse *et al.*, 2014; Guo *et al.*, 2016); *C. ibidis*, which has been detected in feral African Sacred Ibises (*Threskiornis aethiopicus*) in the wetlands of France (Vorimore *et al.*, 2013); *C. suis*, which is thought to be endemic in domestic European pigs and causes enteritis and conjunctivitis (De Puysseleyr *et al.*,

2017); and *C. muridarum*. *C. muridarum* causes murine respiratory tract infections, and was originally classified as a subtype of *C. trachomatis* (Nigg and Eaton, 1944) until 16S and 23S rRNA gene sequencing analysis led to its re-classification as its own species in 1999 (Everett, Bush and Andersen, 1999). *C. muridarum* and *C. trachomatis* share approximately 99% genome sequence identity (Rajaram *et al*., 2015), with most of the genetic differences occurring within the plasticity zone (**1.5.1.1**). Genital tract infection of mice with *C. muridarum* closely mimics acute genital tract infection in women, and consequently *C. muridarum* has been used extensively as a model of chlamydial infection (Barron *et al.*, 1981; Morrison and Caldwell, 2002).

In 1999, a controversial proposal was made to divide the family *Chlamydiaceae* into two genera, *Chlamydia* and *Chlamydophila*, based on sequence differences within the 16S and 23S rRNA genes (Everett, Bush and Andersen, 1999). All species with >90% 16S rRNA identity would be retained in the family *Chlamydiaceae*, and chlamydia-like organisms with 80 - 90% 16S rRNA relatedness to *Chlamydiaceae* would be separated into new families (i.e., *Waddliaceae*, *Simkaniaceae*, and *Parachlamydiaceae*) within the order Chlamydiales. The genus *Chlamydia* was to comprise *C. trachomatis*, *C. muridarum*, and *C. suis*; whilst the genus *Chlamydophila* would consist of *C. pecorum*, *C. pneumoniae*, *C. psittaci*, *C. abortus*, *C. caviae* and *C. felis*. The proposal was widely rejected for many reasons: its over-reliance on minor sequence differences within 16S and 23S rRNA genes; the limited number of isolates of each species included in the analysis; and the new genus designation ignored the conserved biology shared by the organisms that was recognised when they were in a single genus (Schachter *et al.*, 2001; Greub, 2010; Sachse *et al.*, 2015). Further, the proposal stated that to be placed within the same genus, 16S rRNA sequences should be >95% identical. Given that the maximum sequence differences between *C. psittaci*, *C. pneumoniae*, and *C. trachomatis* are between 5 and 6 % (i.e. >94% identity), it was considered unwarranted to separate them into two genera, and such a name change would create confusion. The current, widely accepted classification of the order Chlamydiales is shown in Figure 1, where *Chlamydia* is the sole genus within the family *Chlamydiaceae*, and has multiple species: *C. abortus*, *C. avium, C. caviae, C. felis, C. gallinacea, C. ibidis, C. muridarum, C. pecorum, C. pneumoniae, C. psittaci, C. suis*, and *C. trachomatis* (Stephens *et al.*, 2009).

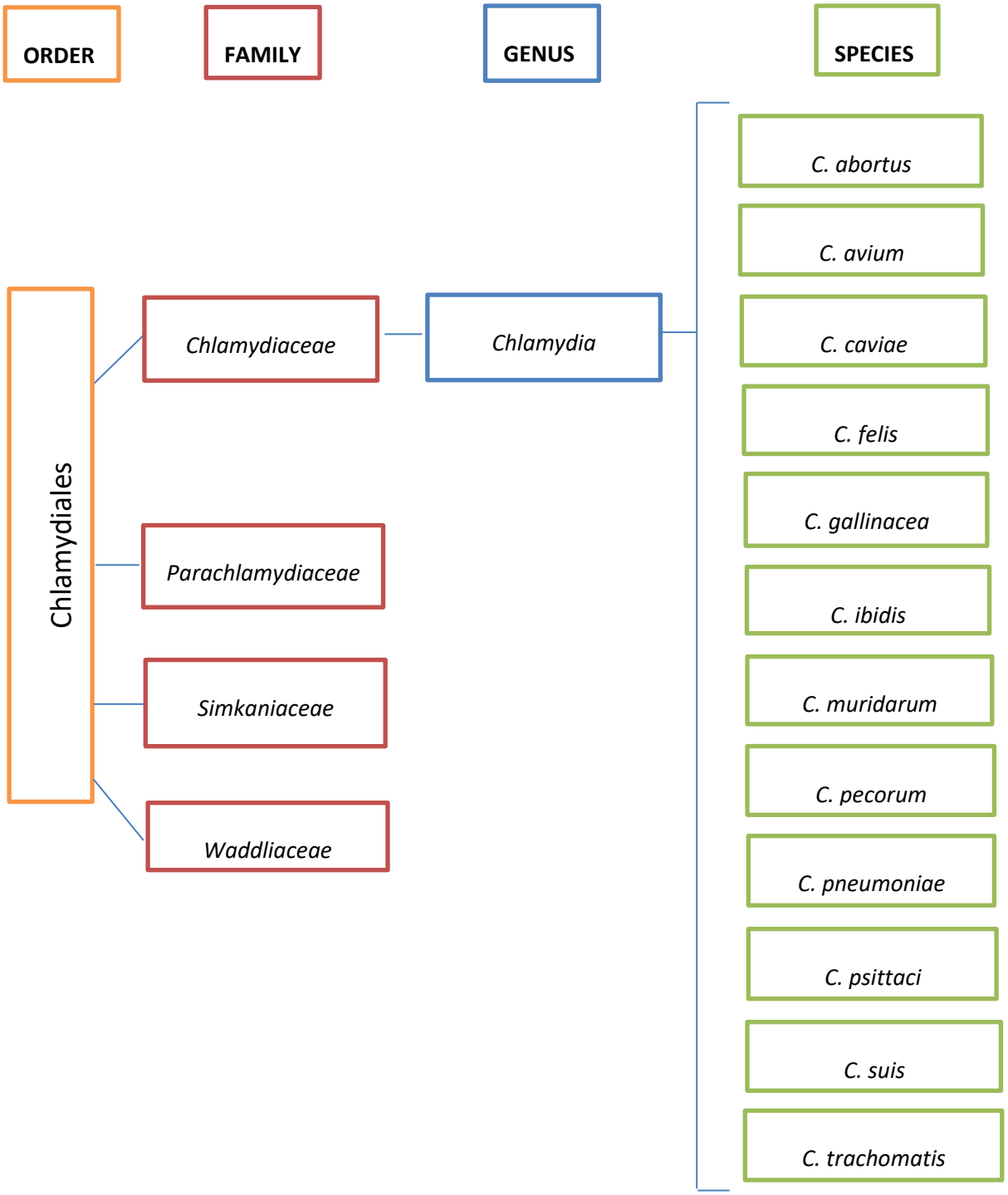


Figure 1 Classification of the order Chlamydiales, resolved to the species level of the genus *Chlamydia*.

Figure was adapted from Vorimore *et al.* (2013), Sachse *et al.* (2015), and Stephens *et al.* (2009).

1.3 Morphology of *C. trachomatis*

C. trachomatis is pleomorphic, existing in two morphological forms during its developmental cycle: the infectious elementary body (EB), and the non-infectious reticulate body (RB), as shown in Figure 2.

EBs are small (~0.3 µm in diameter) and spherical. They contain a highly condensed nucleoid (Abdelrahman and Belland, 2005), and a rigid cell wall that allows for survival of the EB in the extracellular environment. This cell wall contains no peptidoglycan (Moulder, 1991). Its structural rigidity is instead provided by the chlamydial outer membrane complex (COMC); a dense network of heavily cross-linked cysteine-rich proteins including OmcA, OmcB and the major outer membrane protein (MOMP) (Hackstadt, Todd and Caldwell, 1985; Abdelrahman and Belland, 2005). The MOMP is the most abundant protein component of the COMC, accounting for 60% of the mass of the EB outer membrane (Caldwell, Kromhout and Schachter, 1981). It is encoded by the *ompA* gene, and functions as a porin (Bavoil, Ohlin and Schachter, 1984) and as an adhesin (Su *et al.*, 1990) (**1.4** and **1.11**). Also present in the COMC are type III secretion system (T3SS) complexes. The T3SS complex secretes effector proteins into the host cell cytoplasm upon contact of the EB with the host plasma membrane (Elwell, Mirrashidi and Engel, 2016; Hotinger, Pendergrass and May, 2021) (**1.4**). Whilst EBs were initially thought to be metabolically inert, more recent evidence (Omsland *et al.*, 2012; Skipp *et al.*, 2016) has shown that EBs are able to import glucose-6-phosphate (G6P) and store as glycogen or utilise for ATP synthesis (Gehre *et al.*, 2016). It is likely that stores of glycogen in the EB lumen can provide the energy required to sustain the EB during host cell invasion (Cossé, Hayward and Subtil, 2016) or to drive EB differentiation into RBs later in the developmental cycle (Elwell, Mirrashidi and Engel, 2016).

RBs measure approximately 1 µm in diameter. They are adapted for intracellular replication, containing a thin cell wall, and a granular cytoplasm due to ribosome synthesis. RBs are metabolically active and highly express proteins involved in ATP generation and protein synthesis, such as V-type ATP synthases and ribosomal proteins (Elwell, Mirrashidi and Engel, 2016; Skipp *et al.*, 2016).

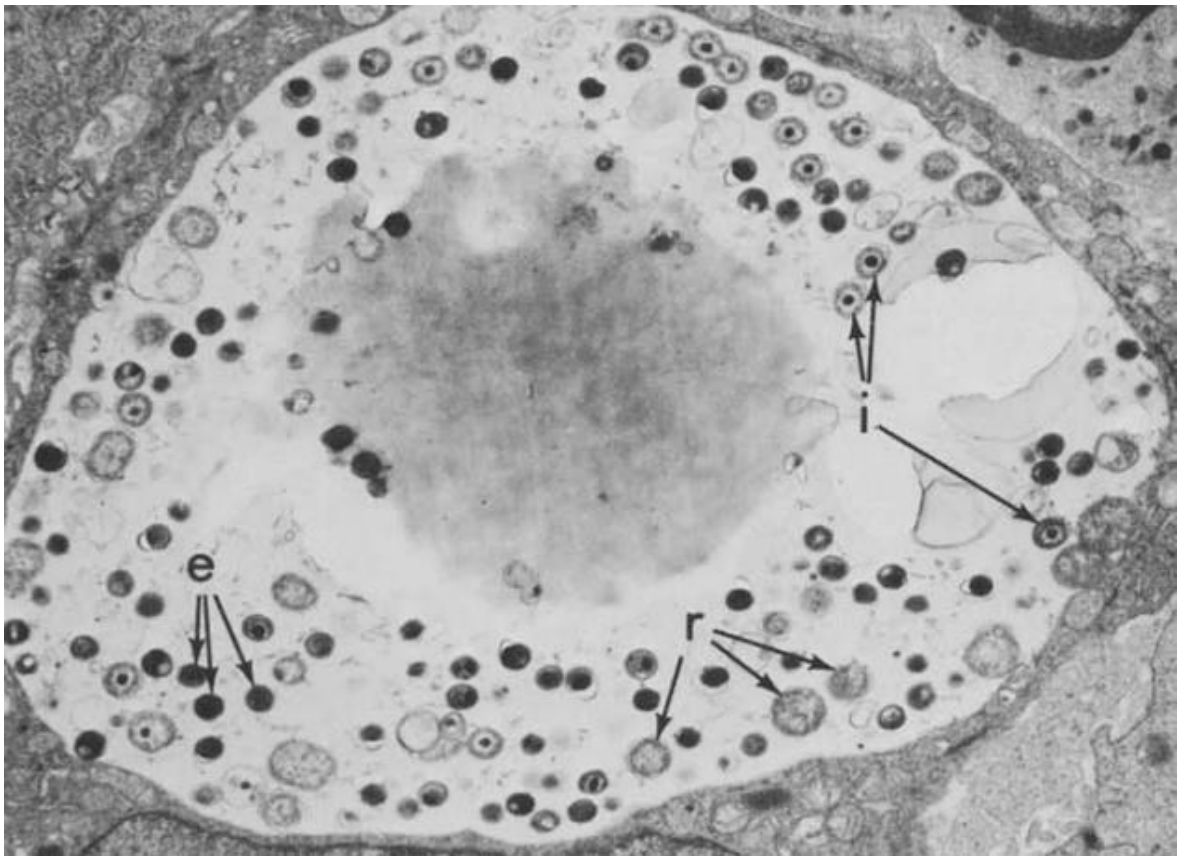


Figure 2 Electron micrograph showing a *C. trachomatis* inclusion (in white) within a human endocervical columnar epithelial cell

The inclusion contains elementary bodies (e), reticulate bodies (r), and intermediate bodies (i). Intermediate bodies are transitional forms of reticulate bodies that are undergoing secondary differentiation to elementary bodies (**1.4**). Magnification is × 12,000**.** Figure taken from Swanson *et al.*, (1975), and reproduced with permission from Oxford University Press.

1.4 Developmental cycle of *C. trachomatis*

C. trachomatis is an obligate intracellular bacterium that primarily infects urogenital tract and ocular columnar epithelial cells. It has a unique biphasic developmental cycle, during which it transitions between two morphological forms – the infectious elementary body (EB), and the replicative, non-infectious reticulate body (RB) **(1.3)**. The developmental cycle of *C. trachomatis* is shown in Figure 3. The first stage of the cycle is the binding of EBs to a host epithelial cell via a trimolecular bridge between host receptors, host heparan sulphate proteoglycans (HSPGs), and adhesins including the major outer membrane protein (MOMP) and outer membrane complex B protein (OmcB) (Su *et al.*, 1996; Stephens *et al.*, 2001; Moelleken and Hegemann, 2008; Cossé, Hayward and Subtil, 2016; Elwell, Mirrashidi and Engel, 2016). Pre-synthesised effector proteins are then injected into the host cell through type III secretion system (T3SS) complexes found on the surface of EBs (**1.3**). This triggers actin re-organisation within the host cell, which facilitates internalisation of the EB into a membrane-bound compartment called an inclusion (Dai and Li, 2014; Elwell *et al.*, 2016). Internalisation into the host cytosol allows *C. trachomatis* to evade detection by the host immune response. The inclusion is transported along microtubules to the microtubule-organising centre (MTOC) of the host cell. The *C. trachomatis* DNA begins to de-condense and gene expression is initiated. Approximately 8 hours post-infection, EBs undergo primary differentiation into RBs (Abdelrahman and Belland, 2005). The RBs lose the thick outer membrane that aided in the survival of the EB in the extracellular environment and divide by binary fission in a rapidly expanding inclusion. RBs typically undergo 8 - 12 rounds of division. Approximately 20 hours post-infection, DNA condenses and RBs undergo secondary differentiation into EBs. A transitional form of the differentiating RB, termed the “intermediate body”, is sometimes visible using electron microscopy at this stage of the developmental cycle (Lee *et al.*, 2018) (**Figure 2**). Under conditions of stress, such as penicillin treatment, nutrient deprivation or interferon-γ (IFN-γ) exposure (**1.5.1.1**); rather than secondary differentiation into EBs, RBs enter a non-dividing persistent state and become enlarged forms termed “aberrant bodies”. Lambden *et al.* (2006) identified that upon addition of penicillin to cultures at 20 hours post-infection, binary fission was blocked and the differentiation from RB to EBs was prevented. However, it was also noted that exposure to penicillin did not prevent chromosomal or plasmid DNA replication. *C. trachomatis*-infected host cells exposed to penicillin did not lyse, but harboured large aberrant RBs in inclusions that filled the cytoplasm. In these inclusions, the DNA replicated beyond the end of the normal chlamydial developmental cycle. At 60 hours post-infection, each RB had a minimum of 16 chromosomal copies. A study by Skilton *et al.* (2009) similarly noted that chromosomal and plasmid DNA replication is unaffected by the addition of penicillin, and identified that after the removal of penicillin, the normal *C. trachomatis* developmental cycle could be recovered within 10 to 20 hours.

In this way, removal of the stress factor reverses the persistent state, resulting in division and differentiation of RBs into EBs (Hogan *et al.*, 2004; Schoborg, 2011; Panzetta, Valdivia and Saka, 2018). EBs are subsequently released into the extracellular environment by host cell lysis or extrusion of the inclusion from the host cell plasma membrane (Hybiske and Stephens, 2007). Released EBs can then infect neighbouring epithelial cells. The *C. trachomatis* developmental cycle is complete after approximately 48 - 72 hours (Miyairi *et al.*, 2006).

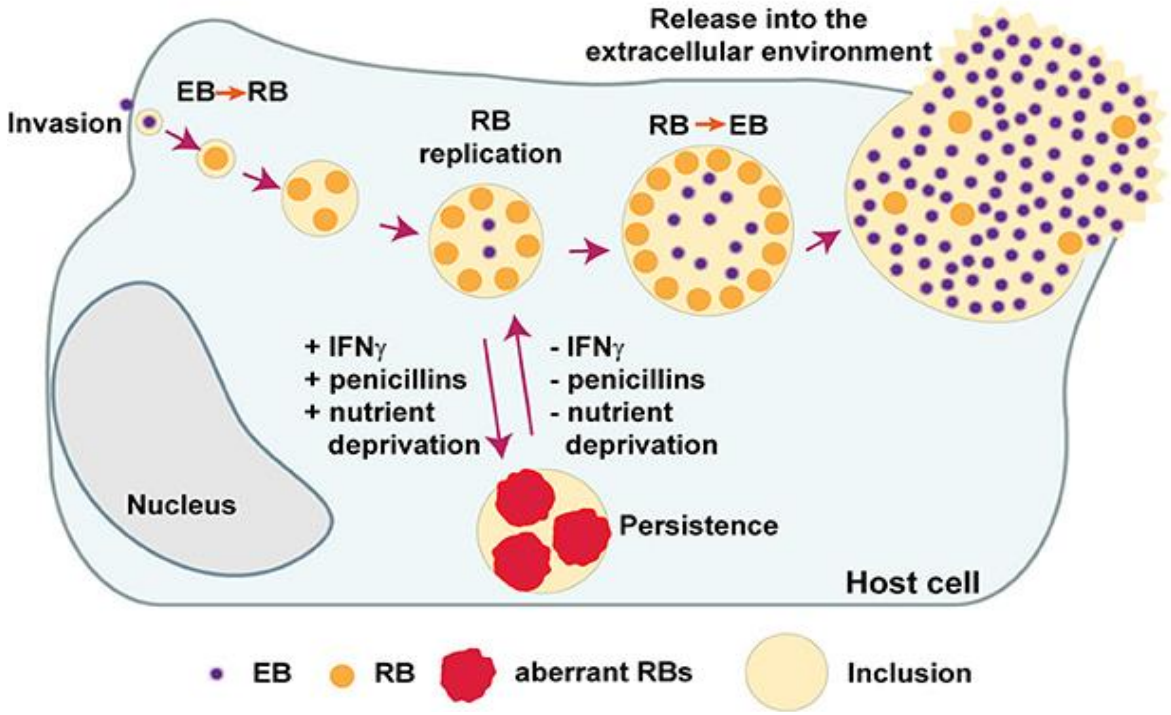


Figure 3 ***C. trachomatis*** **developmental cycle**

Infectious EBs are internalised into host cells in an inclusion. EBs differentiate into RBs that subsequently divide by binary fission. RBs differentiate back into EBs approximately 20 hours post-infection, and EBs are released into the extracellular environment by cell lysis or extrusion of the inclusion, where they can infect neighbouring cells. Under conditions of stress, RBs enter a persistent state and become enlarged ("aberrant bodies"). Removal of the stress factor reverses the persistent state, and RBs divide and differentiate into EBs. Figure was taken from Panzetta *et al.*, 2018 and reproduced under the terms of the Creative Commons 4.0 Licence.

1.5 Biovars and serovars of *C. trachomatis*

C. trachomatis strains are grouped by disease type into biovars. *C. trachomatis* strains comprise two biovars: the trachoma biovar, consisting of strains that cause localised infections of columnar epithelial cells of the conjunctiva or genital mucosa; and the Lymphogranuloma venereum (LGV) biovar, which consists of invasive strains that can spread systemically through the lymphatic system (Brunelle and Sensabaugh, 2006).

C. trachomatis strains have traditionally been classified into serovars using microimmunofluorescence (MIF) (Stephens *et al.*, 1982) (**1.11**). This serological method uses monoclonal antibodies targeted against the MOMP. MOMP, encoded by the *ompA* gene, consists of five conserved domains (CD 1 – 5), and four polymorphic variable domains (VD 1 - 4) that encode surface-exposed epitopes (**1.11.2**). These epitopes are the main targets of the host immune response. The *ompA* gene is subject to host immune selection pressure, and has been identified as a mutational hotspot within the *C. trachomatis* genome (Brunelle and Sensabaugh, 2012) (**1.11.1.3**). Amino acid substitutions within the variable domains result in variation in MOMP epitopes that are serovar-specific.

The two *C. trachomatis* biovars are sub-divided into 15 – 19 serovars: the trachoma biovar includes serovars A – K, of which serovars A – C infect ocular epithelial cells and cause trachoma (**1.6.1**), and serovars D - K that primarily infect epithelial cells of the urogenital tract and cause sexually transmitted urogenital tract infections (**1.6.2**). The LGV biovar consists of serovars L1 - L3, which can infect the lymphatic system and cause Lymphogranuloma venereum (LGV) (**1.6.3**). Serotyping of the *ompA* gene has been superseded by *ompA* genotyping, whereby the *ompA* gene is sequenced to determine the *ompA* genotype (**1.11.3**).

Whole genome sequencing of *C. trachomatis* strains has shown that the species is split into two distinct clades, representing the trachoma and LGV biovars. Of the 1 Mb genome, a total of 4,860 single nucleotide polymorphisms (SNPs) distinguishes the two biovars. The trachoma clade comprises two lineages, T1 and T2. T1 is composed of clinically prevalent urogenital *ompA* genotypes, mostly E and F), and T2 is comprised of less prevalent urogenital *ompA* genotypes. All ocular strains form a cluster within T2, which indicates an emergence from a urogenital ancestor (Harris *et al.*, 2012). Strains within the LGV clade are less diverse than those in the trachoma lineage, with shorter branch lengths (Harris *et al.*, 2012).

1.5.1 Tissue Tropism

There is no functional association between the MOMP and tissue tropism (Seth-Smith *et al.*, 2013; O'Neill *et al.*, 2018); and recombination of the *ompA* gene has been observed between LGV, urogenital and ocular isolates (Harris *et al.*, 2012). It has been suggested that the high level of variability within MOMP allows *C. trachomatis* to evade immune detection (Brunelle and Sensabaugh, 2012; Harris *et al.*, 2012). Other genes within *C. trachomatis* have been implicated as determinants of tissue tropism. These include the translocated actin recruiting phosphoprotein (TarP) (Lutter *et al.*, 2010); the cytotoxin genes (Belland *et al.*, 2001); and members of the polymorphic membrane protein family (*pmp*) (Gomes *et al.*, 2006). The tryptophan synthase operon has also been shown to play a role in defining *C. trachomatis* host tissue tropism (Fehlner-Gardiner *et al.*, 2002; Caldwell *et al.*, 2003). A study by Last *et al.* (2018) applied whole genome sequencing to conjunctival swabs from a treatment-naïve trachoma-endemic population in Guinea-Bissau. The purpose of the study was to identify whether there are any polymorphisms associated with ocular localisation and clinical disease severity. The study identified 21 non-synonymous SNPs associated with ocular localisation, including within the polymorphic membrane protein D (*pmpD)* gene, and *tarP* gene. A total of 8 synonymous SNPs associated with disease severity were identified within the *pmpE*, *glgA*, and *yifH* genes.

1.5.1.1 The Tryptophan Synthase Operon (trpRBA)

The tryptophan synthase operon (*trpRBA*) is closely linked to ocular and genital tropism (Caldwell *et al.*, 2003; O'Neill *et al.*, 2018). Tryptophan synthase is a tetrameric enzyme comprised of two alpha subunits encoded by the *trpA* gene, and two beta subunits encoded by the *trpB* gene. Tryptophan synthase catalyses the two final steps in the biosynthesis of tryptophan, namely the cleavage of indole glycerol 3-phosphate (IGP) to form indole and glyceraldehyde 3-phosphate (the alpha reaction catalysed by TrpA), followed by the conversion of indole and L-serine to L-tryptophan (the beta reaction catalysed by TrpB). X-ray crystallography data has demonstrated that the alpha and beta subunits of tryptophan synthase are connected by an intermolecular hydrophobic tunnel that allows for the diffusion of indole from the alpha reaction in the alpha subunit to the beta subunit for tryptophan biosynthesis (Miles, 2013).

The operon resides in a hypervariable region of the *C. trachomatis* genome termed the plasticity zone (PZ) (Read *et al.*, 2000). Sequencing of the PZ of 15 *C. trachomatis* strains illustrated that 14 of 15 strains encoded homologues of *trpA* and *trpB*. Interestingly, whilst the TrpB protein was highly conserved across all strains, retaining conserved residues essential for enzyme activity (Miles, 2001); the TrpA protein possessed several alterations in its primary structure that differed between

ocular and genital strains of *C. trachomatis* - the *trpA* of ocular but not genital strains of *C. trachomatis* have a common triplet deletion at nucleotide position 408 - 410, that results in the loss of Phe-136 from the TrpA polypeptide. A "mutational hotspot" at nucleotide position 531 was also identified – ocular strains have a single nucleotide deletion that results in a frame shift and subsequent expression of a non-functional and truncated TrpA subunit; whereas genital strains have two point mutations at the same site that result in amino acid substitutions that differ between serovars D - K: CQ in D, E, and K; YQ in G, F, H, I and J; and YE in LGV strains (Caldwell *et al.*, 2003). These amino acids are located in loop 6 of tryptophan synthase, a region that is essential for metabolite channelling (Miles, 2013).

The situation is complicated further in the context of *C. trachomatis* infection in humans. T-cells play an important role in the development of adaptive immunity against *C. trachomatis* infection, and the inflammatory agent interferon-γ (IFN-γ) is a key component of this response (Fehlner-Gardiner *et al.*, 2002), interfering with the ability of *C. trachomatis* RBs to replicate (**1.4**). Upon binding of IFN-γ to its IFN-γ receptor, IFN-γ activates the expression of indoleamine-2,3-dioxygenase (IDO), which degrades L-tryptophan to L-kynurenine. This deprives the *C. trachomatis* RBs of the tryptophan that they require for division and re-differentiation to infectious EBs. *C. trachomatis* deprived of tryptophan are unable to differentiate into EBs, resulting in the loss of infectivity. Interestingly, *Chlamydiae* have seemingly evolved their own rescue system. High levels of IFN-γ inhibit growth of *C. trachomatis*, however, sub-inhibitory concentrations of IFN-γ have been shown to promote a persistent state, whereby RBs form large aberrant bodies (**1.4)** (Beatty et al., 1994). These aberrant bodies have been shown to regain their infectious state after several weeks, following the removal of IFN-γ from the system (Beatty et al., 1995). The ability of *C. trachomatis* RBs to enter this persistent state is a highly advantageous feature, aiding *C. trachomatis* in its avoidance of the host cell immune response. Persistence allows *C. trachomatis* to 'ride out' hostile conditions and maintain a long-term, chronic infection within a host cell. Given that persistence results in the arrest of cell division and temporarily pauses the developmental cycle (Schoborg *et al.,* 2011), this can provide a degree of protection during infection (**1.4**).

As a result of the alterations to TrpA in ocular and genital *C. trachomatis* strains, both genital and ocular strains are unable to convert IGP to indole in the alpha reaction. Importantly, genital but not ocular strains of *C. trachomatis* are rescued under conditions of tryptophan starvation (i.e. IFN-γ presence) by the addition of exogenous indole (Fehlner-Gardiner *et al.*, 2002). It has been suggested that the most likely source of this indole is other colonisers of the female genital tract, such as the *Peptostreptococcus*, *Fusobacterium* and *Bacteroides* species (Gibbs, 1987; Larsen and Monif, 2001). It is thought possible that genital strains have retained the full-length tryptophan synthase (with

mutations) so that they may still be able to synthesise tryptophan in the presence of IFN-γ-mediated host defence mechanisms using this exogenous indole. It was suggested by Caldwell *et al.* (2003) that the ocular environment was sterile and that with no exogenous source of indole (i.e., no other eye colonisers), there was little need to maintain a functional and intact tryptophan synthase. However, the ocular environment is no longer considered sterile. Willcox *et al.,* (2013) noted that swabs of the conjunctiva can yield cultivable microbes, most frequently belonging to the non-indole producing bacterial genera *Propionibacterium*, *Corynebacterium*, and *Staphylococcus*. Garza *et al.,* (2021) also noted a predominance of these bacterial genera within the conjunctiva. A study by Sherchand and Aiyar (2019) suggested that there may be a selection against maintaining an intact tryptophan synthase enzyme in the ocular environment, as, in the absence of indole (such as in the eye), tryptophan synthase deaminates serine to pyruvate and ammonia. The study noted that intact TrpA and TrpB subunits of the tryptophan synthase operon are required for serine deamination, and ammonia is generated when the intact tryptophan synthase enzyme is expressed. It was posited that this ammonia generation may act as a negative selection against maintaining an intact tryptophan synthase in the indole-free ocular environment, which can provide a survival advantage for ocular *C. trachomatis* strains.

1.5.2 **In vitro and in vivo studies of virulence**

Studies of *C. muridarum* have typically used two *C. muridarum* isolates (Weiss and Nigg) interchangeably as these were considered identical and were derived from a single original strain. Research by Ramsey *et al.* (2009) took plaque-purified isolates derived from their stocks and each were whole genome sequenced. The study showed that genomic diversity was present within the two isolates. Specifically, five non-synonymous SNPs were identified in *C. muridarum* Nigg strain when compared to the published *C. muridarum* Nigg strain. In addition, there were a total of 11 SNPs that were shared between the *C. muridarum* Nigg strain under investigation and the *C. muridarum* Weiss strain compared to the published Nigg sequence. Eight of these SNPs resulted in a frame shift and a putative change in protein expression, and nine mutations resulted in a single amino acid change. This research showed that genomic diversity exists within stock isolates of *C. muridarum*. This study by Ramsey *et al.* (2009) also showed that these two strains of *C. muridarum* (Weiss and Nigg) possessed distinct growth characteristics in vitro and varied in their virulence in vivo. The Weiss isolates (n=4) displayed greater virulence in challenge infections. Further, in vitro, the Weiss isolates produced consistently smaller inclusions in HeLa cells, and smaller plaques in monolayers of mouse fibroblast cells compared to the Nigg isolates (n=5). Lastly, the Weiss isolates possessed higher replicative yields in vitro than the Nigg isolates. It was noted by Ramsey *et al.*

(2009) that caution must be applied when ascribing differences in virulence to the genetic variation, as it cannot be asserted with complete confidence that these isolates are representative of the entire population. However, the study by Ramsey *et al.* (2009) did show that there is genotypic and virulence diversity amongst *C. muridarum* isolates.

There are several ways to investigate the impact of virulence beyond a symptomatic or asymptomatic phenotype, using in vitro experimentation. Kari *et al.* (2008) showed that trachoma isolates with the same *ompA* genotype (A2497 and A/Har-13) had a marked difference in virulence in cell culture. In vitro growth rate and plaque morphology, sensitivity to IFN-γ, and infectivity for non-human primates were all investigated in the study. It was shown that A2497 formed large clear plaques five days after infection, whilst A/Har-13 produced small plaques that were undetectable until 11 days after infection. Further, the growth of A/Har-13 was found to be completely inhibited when infected cells were treated with IFN-γ of 2 ng/ml or higher, whereas A2497 was resistant to IFN-γ even at the highest concentration tested (32 ng/ml). Finally, the infectivity of the A2497 and A/Har-13 was compared in a non-human primate model by infecting the eyes of a macaque with either A/Har-13 or A2497. The course of infection was monitored by culture and also by clinical response – the latter by giving a score from 0 (no disease) to 12 (maximal disease) for the appearance of the upper tarsal conjunctiva. The study showed that the average clinical response score was greater for the A2497 strain than for A/Har-13. Kari *et al.* (2008) also measured *C. trachomatis* shedding from infected conjunctivae and showed that A2497-infected monkeys shed ten-fold more infectious organisms than A/Har-13.

A study by Miyairi *et al.* (2011) used a variety of in vitro and in vivo methods to investigate why two *C. psittaci* strains showed a 1000-fold difference in lethal dose in mice. In vitro growth was measured in fibroblasts and peritoneal macrophages however no differences were observed between the two *C. psittaci* strains. The study also considered macrophage activation status and neutrophil recruitment of both strains. It was shown that infection with the less virulent strain led to an influx of activated macrophages with relatively minor organ damage, whereas the more virulent strain resulted in an influx of macrophages and neutrophils and caused significant organ damage.

Rasmussen *et al.* (1997) showed that infection of cervical epithelial cells with *C. trachomatis* can up-regulate mRNA expression and induce secretion of pro-inflammatory cytokines including IL-1α, IL-6, and IL-8. This cytokine response was shown to be sustained throughout the *C. trachomatis* developmental cycle.

1.6 Pathogenicity of *C. trachomatis* in humans

1.6.1 Trachoma

C. trachomatis strains with *ompA* genotypes A - C infect ocular epithelial cells, causing trachoma. Trachoma is the leading infectious cause of blindness, and the most common cause of preventable blindness in the developing world (Weir, Haider and Telio, 2004). An estimated 1.9 million individuals have been blinded by trachoma (World Health Organisation, 2020b) and the disease is endemic in rural and poverty-stricken regions of India, North Africa, South East Asia and Australia. Clinically, trachoma is sub-divided into active (early-stage) and cicatricial (late-stage) disease. Active disease is more commonly found in children aged between 3 - 6 years and is characterised by follicular conjunctivitis, with inflammatory thickening of the conjunctivae in severe cases. Infection is typically acquired through close proximity to those with active disease, and known transmission routes include hand-to-eye contact and the sharing of contaminated items, in addition to eye-seeking flies (Emerson *et al.*, 2000; Burton and Mabey, 2009). A study by Versteeg *et al.* (2020) investigated how long *C. trachomatis* can remain viable at non-ocular sites under lab-controlled conditions. It was suggested that the presence of *C. trachomatis* DNA at non-ocular sites could represent a route of transmission of trachoma. The study determined that plastic, skin, and cotton cloth can facilitate transmission of trachoma-causing *C. trachomatis* strains, by acting as sites where bacterial reservoirs can be deposited and then transferred into previously uninfected eyes.

Cicatricial disease results from repeated infection, which causes chronic conjunctival inflammation and scar tissue formation. The build-up of scar tissue causes the eyelids to turn inwards (entropion) and the in-folded eyelashes to scratch the cornea (trichiasis). This leads to corneal opacity and blindness (World Health Organisation, 1987, 2020b). Cicatricial disease typically occurs from the third decade of life (Burton, 2007) but has been shown to present earlier in highly endemic regions (Emerson *et al.*, 2008; World Health Organisation, 2020b). Whilst active disease can be successfully treated with antibiotics (tetracycline eye ointment or oral azithromycin), treatment of cicatricial disease requires surgery.

The highest prevalence of trachoma is reported in Ethiopia and Sudan, where active disease is present in over 50% of children aged 1 - 9 years, and trichiasis present in 10% of adults (Berhane, Worku and Bejiga, 2006; World Health Organisation, 2020b). In Ethiopia, there appears to be no marked gender difference in the prevalence of active disease, however the prevalence of trichiasis in adult females is two-fold compared to males (4.1% vs 1.6%, p-value <0.001) (Berhane, Worku and Bejiga, 2006). This is likely the result of adult females having increased exposure to infection through contact with children with active disease.

The World Health Organisation (WHO) endorsed the "SAFE" strategy in 1993 with the aim to eliminate trachoma globally by 2020: Surgery to correct eyelid deformities, Antibiotics for individuals with active disease and mass drug administration (MDA) to at-risk populations to reduce disease burden, promotion of Facial cleanliness, and Environmental improvements (i.e. provision of water and improved sanitation) (World Health Organization, 1997). As of September 2020, ten countries have been validated by the WHO as having eliminated trachoma as a public health problem. In 2019, 92,622 people received corrective surgery for trichiasis, and 95.2 million individuals in endemic communities were treated with antibiotics (World Health Organisation, 2020b). Despite these strides, 137 million people still live in trachoma-endemic areas, and trachoma remains a public health problem in 44 countries (World Health Organisation, 2020b). As a result, the global trachoma elimination target was not reached in 2020. The new target date for global trachoma elimination is set for 2030 (World Health Organisation, 2020a).

The progress made towards global trachoma elimination could be threatened by the increased displacement of humans from trachoma endemic regions. Of the 68.5 million people categorised as forcibly displaced in 2017 (United Nations High Commissioner for, 2019), over 50% are from countries where trachoma is considered endemic, including South Sudan. Population-based surveys of trachoma within refugee camps serving South Sudanese refugees in Sudan (Sanders *et al.*, 2019) have shown that the prevalence of active disease amongst children aged 1 - 9 years in the camps was greater than 10% and the prevalence of trichiasis in those aged 15 years and above was greater than 2.9%, highlighting that trachoma is prevalent amongst those living in the refugee camp setting. To have any hope of eliminating trachoma, these displaced populations must also be provided with comprehensive SAFE interventions.

1.6.2 **Urogenital Tract Infections**

C. trachomatis strains with *ompA* genotypes D - K primarily infect columnar epithelial cells of the urogenital tract, causing urogenital tract infections that are transmitted during unprotected oral, vaginal, or anal sex. Urogenital *C. trachomatis* infections can also be transmitted from infected mother to neonate during childbirth, resulting in Opthalmia Neonatorum (neonatal conjunctivitis) or pneumonia in some exposed infants (Darville, 2005; Zar, 2005b) (**1.7**). Urogenital *C. trachomatis* infections are asymptomatic in approximately 50% of males and 75% of females (Stamm, 1999; Lanjouw *et al.*, 2016). This delays diagnosis and facilitates un-interrupted *C. trachomatis* transmission. Those with symptoms report pain during urination, abnormal penile or vaginal discharge, and painful sexual intercourse. Urogenital *C. trachomatis* infections are typically treated with a single dose of azithromycin. If left untreated, urogenital *C. trachomatis* infections can ascend

the urogenital tract and cause serious sequelae (Hoenderboom *et al.*, 2019), including pelvic inflammatory disease (PID), which is characterised by chronic inflammation of the fallopian tubes. In men, epididymitis can result from un-treated urogenital *C. trachomatis* infections.

C. trachomatis strains with *ompA* genotypes D - K can also infect columnar epithelial cells of the rectum (Hamlyn and Taylor, 2006), and genotypes D, G and J are frequently detected in rectal swabs collected from men who have sex with men (MSM) engaging in receptive anal intercourse (Geisler *et al.*, 2002; Quint *et al.*, 2011). Whilst two-thirds of rectal non-LGV infections are asymptomatic (Hamlyn and Taylor, 2006), some individuals experience mild proctitis. Rectal non-LGV *C. trachomatis* infections are treated with doxycycline (100 mg) twice daily for a minimum of 7 days.

C. trachomatis strains with *ompA* genotypes D – K can also cause adult inclusion conjunctivitis (Kestelyn, 2008). Adult inclusion conjunctivitis is sexually transmitted and occurs most frequently in sexually active young adults. The disease is transmitted primarily through hand-to-eye spread of infected genital secretions. Symptoms include unilateral mucopurulent discharge, swollen eyelids and crusting of eyelashes. Many patients (up to 61% of males and 90% of females) with adult inclusion conjunctivitis have a concurrent urogenital *C. trachomatis* infection (Mitchell *et al.*, 2021). Patients and their sexual partners are treated with systemic antibiotics, typically a single dose of azithromycin (Lanjouw *et al.*, 2016).

Extracellular survival and transmission of *C. trachomatis* has been considered in numerous studies, including those by Giffard *et al.* (2018), Lewis et al. (2012), and Janssen *et al.* (2020). Giffard *et al.* (2018) demonstrated that human fingers contaminated with *C. trachomatis* can contaminate urine specimens by contact. The study identified that three contaminated fingers dried on a paper towel transfers approximately 10 µL of *C. trachomatis* suspension to a urine specimen. Similarly, Lewis *et al.* (2012) showed that there is the potential for contamination of containers used to collect self-taken swabs owing to surface contamination with *C. trachomatis* DNA within a sexual health clinic. A study by Janssen *et al.* (2020) demonstrated that viable *C. trachomatis* could be detected from rectal swabs collected from *C. trachomatis*-diagnosed women, irrespective of reported anal sex or rectal symptoms. The findings suggested that rectal *C. trachomatis* detected in some women may represent a currently on-going infection rather than the presence of remnant DNA from dead bacteria or contamination from an active vaginal *C. trachomatis* infection.

1.6.3 **Lymphogranuloma venereum (LGV)**

C. trachomatis strains with *ompA* genotypes L1 - L3 cause Lymphogranuloma Venereum (LGV), an invasive sexually transmitted disease. The clinical course of LGV is divided into three stages: local

infection (primary stage), regional dissemination (secondary stage), and progressive tissue damage (tertiary stage). *C. trachomatis* enters through breaks in the skin, and infects monocytes and macrophages. This results in localised inflammation at the inoculation site, which causes a painless, self-resolving ulcer to form. Approximately 2 - 6 weeks after the initial appearance of the ulcer, *C. trachomatis* disseminates to local draining lymph nodes (Fehlner-Gardiner *et al.*, 2002; Mabey, 2002). This results in lymph node enlargement near the infection site (lymphadenitis), and inflammation of the walls of lymphatic vessels (lymphangitis). The clinical presentation at this stage depends on the inoculation site – inoculation at the mucosal lining of the penis or vagina presents as an inguinal syndrome characterised by swelling of the inguinal lymph nodes to form buboes; whereas inoculation at the mucosal lining of the rectum presents as an anorectal syndrome characterised by proctitis and lower abdominal pain. The tertiary stage occurs if LGV is left untreated: chronic inguinal lymphangitis leads to fibrosis, which results in lymphatic obstruction, causing lymphedema and elephantiasis of the genitals. Tertiary complications of anorectal LGV include rectal strictures and perirectal abscesses (O'Byrne *et al.*, 2016).

LGV was originally considered a tropical disease, confined to heterosexual populations of resource-poor countries (Mabey, 2002). Sporadic cases were reported in the western world, however these were mainly imports from endemic regions. This largely remained the case until 2003, when an outbreak of LGV was reported in the Netherlands amongst MSM presenting with proctitis (Nieuwenhuis *et al.*, 2003; Götz *et al.*, 2004) (**1.10**). Sequencing of the *ompA* gene demonstrated that the outbreak was caused by a new L2 variant, designated L2b (Spaargaren, Fennema, *et al.*, 2005). This variant has subsequently been implicated in LGV outbreaks worldwide, including within the United Kingdom (Nieuwenhuis *et al.*, 2004; Kropp, Wong and Canadian LGV Working Group, 2005; Spaargaren, Fennema, *et al.*, 2005; Morton *et al.*, 2006; Ward *et al.*, 2007). Between 2003 and 2018, there have been 6,752 UK LGV diagnoses (Public Health England, 2017; European Centre for Disease Control, 2019). The LGV epidemic in MSM in the United Kingdom is discussed in section 1.10.

1.7 Neonatal *C. trachomatis* infections

Pregnant females with urogenital chlamydial infections can transmit *C. trachomatis* to their infants during vaginal delivery, when the infant becomes coated with *C. trachomatis* as it passes along the birthing canal. The estimated prevalence of *C. trachomatis* in pregnant women in the UK is 1.96 – 2.4% (Oakeshott *et al.*, 2002; Shankar *et al.*, 2006; Kirk *et al.*, 2008). Between 50 - 70% of infants born to infected mothers acquire *C. trachomatis* if prophylaxis is not provided before or immediately following delivery (Joseph Davey, Medline and Klausner, 2016). Infected infants can develop conjunctivitis (Ophthalmia Neonatorum). Neonatal conjunctivitis typically develops 5 - 14 days after birth and is characterised by eyelid oedema and ocular discharge.

Nasopharyngeal *C. trachomatis* infections or aspiration of infected genital secretions during delivery, can cause pneumonia in newborns. Pneumonia affects approximately 10 - 20% of infected infants and occurs typically 1 - 19 weeks after birth (Hammerschlag, 2011). Pneumonia is characterised by a low-grade fever and persistent cough. If left un-treated, infants are at increased risk of developing chronic pulmonary disease (Zar, 2005a). The recommended treatment for neonatal chlamydial conjunctivitis and pneumonia is oral erythromycin for 14 days (Nwokolo *et al.*, 2016). In the UK, chlamydia screening is not provided as part of routine antenatal care, however, health professionals are recommended to inform pregnant women aged <25 years about the higher prevalence of *C. trachomatis* within their age group (**1.8.1**) and to provide details on screening.

1.8 High-risk groups for *C. trachomatis* infection

There are several groups of individuals at high risk of becoming infected with *C. trachomatis*. These include the following:

1.8.1 Individuals aged 15 to 24 years

Younger age is associated with increased risk of *C. trachomatis* infection. The highest rates of *C. trachomatis* infections are amongst sexually active young adults aged 15 to 24 (**1.9**) (**Figure 5**). This is largely due to differences in sexual behaviour within this age-group. According to a national survey of sexual behaviour, those aged 16 to 24 were the most likely of all age groups surveyed to report two or more sexual partners of the opposite sex and at least one new sexual partner of the opposite sex, in the previous year (Mercer *et al.*, 2013). A rapid turnover of sexual partners is associated with higher infection rates of *C. trachomatis* due to the increased likelihood of sexual intercourse with an infected partner (Navarro *et al.*, 2002). The survey (Mercer *et al.*, 2013) also identified that those aged 16 to 24 reported the highest frequency of un-protected sexual intercourse of all age-groups surveyed. Un-protected sex increases *C. trachomatis* infection risk (Navarro *et al.*, 2002) as it allows un-hindered *C. trachomatis* transmission across mucosal epithelia.

Young females are at increased risk of *C. trachomatis* infection compared to their male counterparts (**Figure 5**). In England in 2019, females aged 15 to 24 accounted for 38% of all new *C. trachomatis* diagnoses (87,234/229,411) (Public Health England, 2020b). Higher rates of infection amongst young females can be attributed to several factors. Firstly, older females may have acquired partial immunity to *C. trachomatis* from prior infections (Navarro *et al.*, 2002). Cervical ectopy may also play a role: during puberty, the endocervix (comprised of columnar epithelium) can evert due to hormonal changes, exposing endocervical columnar cells on the ectocervix (Jacobson *et al.*, 2000). This can increase susceptibility to urogenital *C. trachomatis* infection, as *C. trachomatis* strains with *ompA* genotypes D – K primarily infect this cell type (**1.6.2**). Higher rates of infection amongst adolescent females may also be due to greater uptake of chlamydia screening through the National Chlamydia Screening Programme (NCSP) (Public Health England, 2019), which provides opportunistic screening to 15 to 24 year olds - resulting in increased detection of *C. trachomatis* infections. Overall, the prevalence of *C. trachomatis* infections is higher in females of all age-groups than in males (**Figure 5**), and in 2019, 11% more females were diagnosed with *C. trachomatis* infections than males in England (Public Health England, 2020b).

1.8.2 Minority Ethnic Groups

The highest population rates of *C. trachomatis* diagnoses are amongst individuals of black ethnicity, although this varies amongst Black Minority Ethnic (BME) groups (Public Health England, 2021). Individuals of black Caribbean and Black African ethnicity have the highest *C. trachomatis* diagnostic rates of all ethnic groups. Higher rates of infection in these populations likely reflect differences in sexual risk behaviour. Males of black Caribbean and black African ethnicity report higher numbers of recent sexual partners and concurrent partnerships than any other ethnic group (Wayal *et al.*, 2017), which can facilitate *C. trachomatis* transmission within these populations. Ethnic mixing patterns may also play a role: assortative sexual mixing, whereby sexual partnerships are formed between individuals with the same ethnic background, is common within the Black Caribbean and Black population and may assist in the spread of *C. trachomatis* within this sub-population (Aral, 2000).

1.8.3 Men who have Sex with Men (MSM)

Men who have sex with men (MSM) are at high risk of *C. trachomatis* infection, and diagnoses of *C. trachomatis* within this group have risen steadily over the past decade (Public Health England, 2020c). This may be due to increased detection of *C. trachomatis* within MSM (Public Health England, 2020c), however it may also be driven by behavioural changes within this group, including increases in partner numbers, 'chem-sex' (sex using drugs to enhance sexual experience), and the increasing popularity of geosocial networking applications (i.e. Grindr) which facilitate casual sexual encounters and group sex (Gilbart *et al.*, 2013, 2015). Chem-sex has been shown to promote dis-inhibitive behaviour including condom-less sex, and willingness to seek new sexual experiences including fisting and scat play (Gilbart *et al.*, 2013). This increases the risk of *C. trachomatis* acquisition and transmission. MSM report higher numbers of casual and regular sexual partners than heterosexual populations, and are more likely to belong to dense sexual networks (McDaid and Hart, 2010). This enables rapid *C. trachomatis* transmission within the MSM population.

Lymphogranuloma venereum (LGV) is diagnosed almost exclusively amongst MSM (**1.10**) (Public Health England, 2020d). The diagnostic rate of LGV is highest amongst HIV-positive MSM (Public Health England, 2020d). A high LGV incidence within this population is likely due to seroadaptive behaviours including serosorting. Serosorting, whereby HIV-positive MSM can preferentially select other HIV-positive MSM for un-protected sex (Ward *et al.*, 2007), increases exposure risk to other STIs including LGV and hepatitis C virus (HCV). Approximately 20% of LGV infections are reported by MSM with HCV co-infections (Hughes *et al.*, 2013). Since 2017, there has been a rise in LGV cases

amongst HIV-negative MSM; as of 2018, more LGV diagnoses were made in HIV-negative MSM than HIV-positive MSM (Public Health England, 2020a). This could be linked to the increasing availability of pre-exposure prophylaxis (PrEP). There is evidence that PrEP uptake in HIV-negative MSM can result in altered sexual risk behaviour (Traeger *et al.*, 2019), including increased frequency of condom-less sex; which can enhance transmission risk of other sexually transmitted infections (STIs) including LGV (**1.10**).

1.9 Epidemiology of *C. trachomatis* in England

C. trachomatis is the most commonly diagnosed bacterial STI worldwide, with an estimated 131 million new infections reported annually (World Health Organization, 2016). In 2019, chlamydia accounted for 49% of all new STI diagnoses made in England (229,411/468,342) (Public Health England, 2020c). This figure is likely an under-estimate as *C. trachomatis* infections are frequently asymptomatic (**1.6.2**), hence many cases are left un-diagnosed (Stamm, 1999).

The number of *C. trachomatis* diagnoses in England has steadily increased in recent years, rising from 202,966 in 2015 to 229,411 in 2019 (**Figure 4**). During the same time period, diagnoses of other STIs also increased in England: gonorrhoea (from 41,382 in 2015 to 70,936 in 2019), syphilis (from 5,323 to 7,982) and anogenital herpes (from 33,962 to 34,570) (**Figure 4**). This rise in STI diagnoses may be linked to increased STI screening: between 2015 and 2019, there was a 31% increase in sexual health screens across England, from 1,657,425 in 2015 to 2,175,525 in 2019 (Public Health England, 2020c). The National Chlamydia Screening Programme (NCSP) in England provides *C. trachomatis* screening to sexually active young adults aged 15 to 24 (**1.8.1**). The aims of the NCSP are to prevent onward *C. trachomatis* transmission and to reduce prevalence of associated sequelae, through early detection and treatment. In 2019, over 1.3 million *C. trachomatis* tests were carried out amongst 15 to 24 year olds, with a total of 135,167 *C. trachomatis* diagnoses made within this group (Public Health England, 2020c). Young adults, specifically females aged 15 to 24, account for the highest diagnostic rate of *C. trachomatis* of all age groups (**Figure 5**) and possible reasons for this were discussed in **1.8.1**. Figure 5 also shows the increase in diagnoses of *C. trachomatis* across all other age-groups between 2015 and 2019 (except the 13 - 14 year age-group, in which diagnoses decreased between 2015 and 2019) (**Figure 5**).

Individuals of Black ethnicity, particularly Black Caribbean, have the highest population rate of *C. trachomatis* diagnoses of all ethnic groups (Furegato *et al.*, 2016). In 2019, the diagnostic rate of *C. trachomati*s in Black individuals (1,188.6 per 100,000 population) was four times higher than that in White individuals (290.5 per 100,000 population), and seven times higher than in Asian

individuals (168.6 per 100,000 population) (Public Health England, 2020b). The higher diagnostic rate within Black individuals may reflect the differences in sexual risk behaviour within this population.

In recent years, there has been a substantial increase in the number of new *C. trachomatis* diagnoses amongst men who have sex with men (MSM). The diagnostic rate of *C. trachomatis* within this population has almost doubled between 2015 and 2019 (from 2,277.7 to 4,162.8 per 100,000 population). This may be linked to behavioural changes within MSM including increasing use of 'chem-sex' drugs and mobile applications to meet sexual partners, or increased rates of partner change and un-protected sex (McDaid and Hart, 2010; Gilbart *et al.*, 2015). Risk behaviours of MSM are explored further in **1.8.3**.

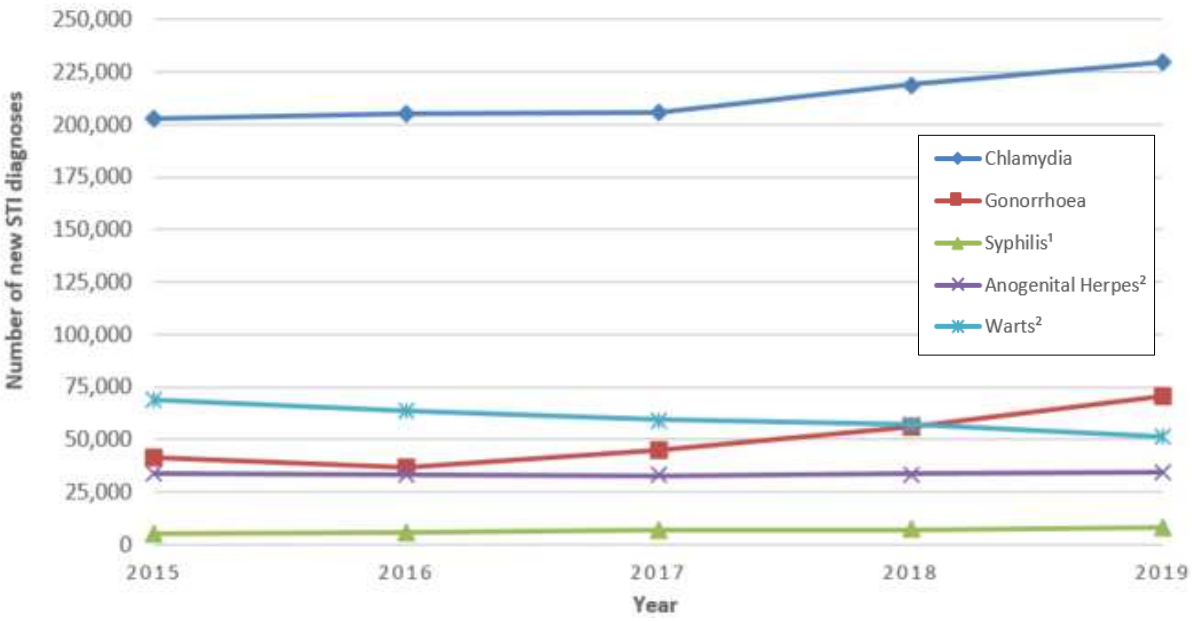


Figure 4 Number of new diagnoses of selected sexually transmitted infections (STIs) made at sexual health services (SHS)† in England between 2015 and 2019. Data taken and adapted from (Public Health England, 2020b). †Sexual health services (SHS) refer to specialist SHS including genitourinary medicine (GUM) clinics; and non-specialist SHS including general practice, pharmacies, and other community-based settings. [1] primary, secondary, and early latent stages; [2] first episode.

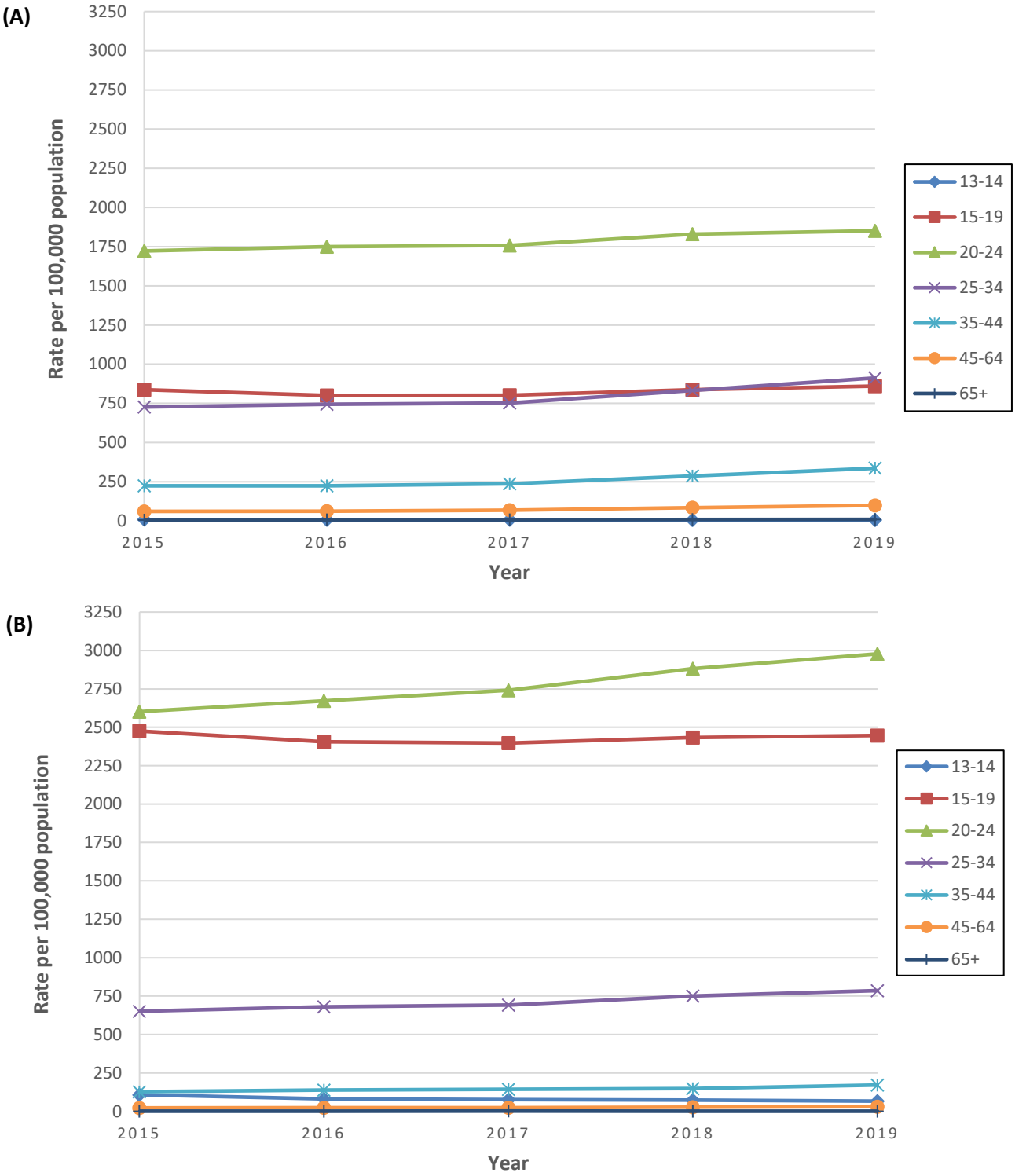


Figure 5 Rate of new urogenital *C. trachomatis* diagnoses made at sexual health services† in England by age group in (A) males and (B) females, 2015 - 2019. Rate given per 100,000 population. Data taken and adapted from (Public Health England, 2020b).

† Sexual health services (SHS) refers to specialist SHS including genitourinary medicine (GUM) clinics; and non-specialist SHS including general practice, pharmacies and other community-based settings.

1.10 The Lymphogranuloma venereum (LGV) epidemic in MSM

1.10.1 A history of the LGV epidemic in MSM

Prior to 2003, Lymphogranuloma venereum (LGV) was considered a tropical disease, largely confined to heterosexual populations in Africa, South East Asia, and the Caribbean (**1.6.3**). Within these populations, the disease manifested as an inguinal syndrome (Ceovic and Gulin, 2015) (**1.6.3**). In the western world, LGV cases were sporadic and most were considered to be imported, affecting travellers from endemic regions or those with sexual contacts from endemic regions (Scieux *et al.*, 1989; Ceovic and Gulin, 2015). However, three clusters of LGV were reported in industrialised, non-tropical regions amongst men who have sex with men (MSM) between 1980 and 2003 (Richardson and Goldmeier, 2007). In Seattle, three MSM were diagnosed with LGV in 1980 after presenting in clinic with rectal symptoms (Levine, Smith and Brugge, 1980), and a further twelve cases were identified in MSM with proctitis in Seattle between 1981 and 1991 (Bauwens *et al.*, 1995). In Paris, 27 MSM were diagnosed with LGV between 1983 and 1989 (Scieux *et al.*, 1989). Of these 27 cases, 25 presented in clinic with inguinal lymphadenopathy, and 2 presented with proctitis (Scieux *et al.*, 1989).

In February 2003, a HIV-positive white bisexual male was diagnosed with LGV in a Rotterdam STI clinic after presenting with bilateral swollen inguinal lymph nodes and a perianal ulcer (Nieuwenhuis *et al.*, 2003). The male had engaged in unprotected receptive anal intercourse with another white male in Amsterdam three weeks prior to presentation at the STI clinic. Genotyping of the *ompA* gene encoding the *C. trachomatis* major outer membrane protein (MOMP) (Morré *et al.*, 1998) (**1.11**) identified *C. trachomatis ompA* genotype L2 in both urine and rectal specimens obtained from the patient.

In April 2003, a further two HIV-positive MSM were diagnosed with LGV in Rotterdam. The *C. trachomatis ompA* genotype for both patients was L2. Both patients presented with proctitis and had no inguinal lymphadenopathy. Contact tracing of 12 recent sexual partners identified 54 men that formed an international sexual network of MSM (Nieuwenhuis *et al.*, 2004). Of these, 13 MSM were diagnosed with LGV. No patients reported an inguinal syndrome, however 12 (92%) reported intestinal symptoms. Genotyping results were provided for nine of the 13 MSM, of which eight were *ompA* genotype L2 and one was *ompA* genotype L1. Of the 13 MSM diagnosed with LGV, 11 (85%) were HIV-positive and 6 (46%) were co-infected with an additional STI including gonorrhoea (n= 4; 31%), and syphilis (n= 1; 8%).

In January 2004, the European Surveillance network of STIs (ESSTI) issued an international alert relating to the LGV outbreak in Rotterdam. By late 2004, cases of LGV proctitis had been reported in Paris (Halioua *et al.*, 2004), Hamburg (Robert Koch-Institut, 2004), and Antwerp (Vandenbruaene, 2004). In October 2004, the UK Health Protection Agency (HPA, now Public Health England) set up an enhanced LGV surveillance system consisting of a reference laboratory diagnostic service and a case reporting system. An alert was issued to clinicians in GUM clinics and microbiologists briefing on the LGV cases across Europe. A leaflet was distributed by the Terrence Higgins Trust warning MSM about LGV and describing its symptoms. By December 2004, 5 cases of LGV proctitis had been reported in the UK and a national incident team was established that included epidemiologists, microbiologists, and clinicians. Clinicians were advised to test all MSM presenting at GUM clinics with symptoms of proctitis or inguinal lymphadenopathy for *C. trachomatis*. By January 2005, a total of 24 cases of LGV were confirmed by *ompA* genotyping in the UK (Macdonald *et al.*, 2005). All cases were the L2 *ompA* genotype. Surveillance data was available for 19 of the patients: all were MSM; 17 (89%) were HIV-positive; 18 (95%) reported anorectal symptoms and 2 (11%) had swollen inguinal lymph nodes.

In July 2005, the results of a retrospective study of 74 MSM diagnosed with rectal *C. trachomatis* infections at an Amsterdam STI clinic between January 2002 and December 2003, were published (Spaargaren, Fennema, *et al.*, 2005). In total, 45 patients were infected with LGV strains of *C. trachomatis*, and sequencing of the *ompA* gene (Morré *et al.*, 1998) revealed that all 45 patients contained a new LGV *ompA* variant called L2b. The nucleotide sequence differences that distinguish the L2b variant from *ompA* genotype L2 are detailed in **1.11**. Of the 45 patients with the new variant, 32 (71%) had presented at clinic with proctitis and 13 (29%) were asymptomatic. Thirty one patients (69%) were HIV-positive. To provide a rough estimate of how long the new variant had been circulating within MSM populations, the *ompA* sequences of *C. trachomatis*-positive rectal swabs from MSM attending the same Amsterdam clinic between 2000 and 2001, and archived swabs from MSM in San Francisco from 1979 to 1985, were compared to the *ompA* sequence of the newly identified L2b variant (Spaargaren, Schachter, *et al.*, 2005). In total, 2 of 67 samples from Amsterdam in 2000, and 4 of 28 samples from 2001 were LGV-positive. All six of these samples contained the L2b variant, which suggested that the variant had been present in MSM in Amsterdam as early as 2000. *OmpA* sequencing of the archived LGV-positive specimens from San Francisco between 1979 and 1985 detected 15 specimens as L1, 18 as L2, and 18 as the L2b variant. These 18 L2b specimens had an identical nucleotide sequence to the Amsterdam L2b variant. This suggested that the LGV outbreak in Europe had most likely been slowly evolving over at least three decades, and could have originated in North America (Spaargaren, Schachter, *et al.*, 2005). The L2b

variant has subsequently been implicated in outbreaks of LGV characterised by proctitis worldwide (Nieuwenhuis *et al.*, 2004; Kropp, Wong and Canadian LGV Working Group, 2005; Ward *et al.*, 2007; Rodriguez-Dominguez *et al.*, 2015; Isaksson *et al.*, 2017).

1.10.2 **Transmission Routes and Clinical Manifestation of LGV in MSM**

LGV strains of *C. trachomatis* are transmitted by direct inoculation of the rectal mucosa during un-protected anal sexual intercourse (de Vries *et al.*, 2019). Ano-oral sex can act as a secondary transmission route for LGV, resulting in inoculation of the pharynx. Pharyngeal LGV infections are rare but can manifest as cervical lymphadenopathy (Korhonen *et al.*, 2012; Dosekun *et al.*, 2013). Genital infections are rarely seen in MSM (de Vrieze *et al.*, 2017). The primary manifestation of LGV in MSM is proctitis, defined as an inflammatory syndrome of the rectal lining. LGV proctitis is characterised by severe anorectal pain, rectal bleeding and haemopurulent discharge. These symptoms resemble, and may be mis-diagnosed as, inflammatory bowel disease (IBD) (Soni *et al.*, 2010; Di Altobrando *et al.*, 2019). Between 27% - 43% of LGV infections are asymptomatic (Spaargaren, Fennema, *et al.*, 2005; Saxon *et al.*, 2016). This is concerning given that asymptomatic individuals may remain un-diagnosed, allowing un-interrupted transmission of LGV *C. trachomatis* strains within sexual networks.

1.10.3 **Diagnosis of LGV**

In the UK, a two-step procedure is followed to diagnose LGV. The first step is to screen clinical specimens (including rectal swabs, pharyngeal swabs, ulcerous material from anogenital lesions, and urine) for *C. trachomatis* using a *C. trachomatis*-specific nucleic acid amplification test (NAAT). This test does not allow strain discrimination, necessitating a second diagnostic step (Ceovic and Gulin, 2015). If specimens test positive for *C. trachomatis*, they are subsequently tested for LGV using an LGV-specific NAAT. This NAAT is a triplex real-time PCR assay that targets a 36 bp deletion in the chromosomal polymorphic membrane protein H gene (*pmpH*) of LGV strains; in addition to an 88 bp region of the *C. trachomatis* plasmid; and the human RNase P gene as an internal process control. This triplex LGV PCR assay was developed by the Sexually Transmitted Bacteria Reference Laboratory (STBRL) at Public Health England (Morré *et al.*, 2005; Alexander, Martin and Ison, 2007; Chen *et al.*, 2007). In the UK, there have been several changes to national guidelines for LGV testing in recent years (Public Health England, 2020d). As of 2015, all individuals with LGV symptoms, contacts of confirmed LGV cases, and all HIV-positive MSM that test positive for *C. trachomatis* at any site, are tested for LGV (Nwokolo *et al.*, 2016).

1.10.4 Treatment of LGV

The recommended treatment for LGV is doxycycline (100 mg) taken orally twice daily for 21 days. This treatment is also recommended for asymptomatic patients and sexual contacts of LGV patients, in addition to patients with inguinal LGV (de Vries *et al.*, 2019). There is no evidence that a HIV co-infection requires a different therapeutic approach (de Vries *et al.*, 2019). There are no guidelines for treating pharyngeal LGV infections, however 7 day, 14 day and 21 day courses of doxycycline (100 mg) have all been shown to be effective in resolving infection (Dosekun *et al.*, 2013). Patients with complications of anorectal LGV (**1.6.3**) including rectal strictures may require surgical intervention in addition to antibiotic treatment.

1.10.5 Epidemiology of the LGV Epidemic in MSM in the UK

1.10.5.1 Case Reporting of LGV

In the UK, the first LGV case reporting system was introduced by the Health Protection Agency (HPA; now Public Health England) in 2004 to monitor the growing LGV epidemic in MSM (**1.10.1**). For each confirmed case of LGV, the HPA sent a questionnaire to be filled out by the clinician that included patient demographic information, sexual history, sexual behaviour, symptoms, other STI diagnoses, and treatment regimen (Ward *et al.*, 2007). Since 2011, all LGV diagnoses made at sexual health services across England are reported to an electronic routine surveillance database: the Genitourinary Medicine Clinic Activity Dataset (GUMCAD), which is operated by Public Health England (Savage *et al.*, 2014). GUMCAD also records laboratory STI tests and patient demographic data including age, sexual orientation, ethnicity, HIV status, and STI history.

1.10.5.2 LGV diagnoses made in England: 2004 – 2019

Between 2004 and 2018, 6,637 LGV diagnoses were made in the United Kingdom. The United Kingdom reports the highest number of confirmed LGV cases of any European country (European Centre for Disease Control, 2019). This may be impacted by surveillance bias, as the UK has had a comprehensive LGV surveillance system in place since 2004 (**1.10.5.1**).

Between 2018 and 2019, there was a 28% increase in LGV diagnoses in England (from 791 to 1,094). Prior to 2019, the number of LGV diagnoses made in England peaked in 2015, then declined until 2017, and has since risen rapidly (Public Health England, 2020d). It is possible that this increase in diagnoses may be the result of increased screening in MSM (Public Health England, 2020d). Between 2011 and 2019, the annual number of LGV tests performed in England increased steadily, and this could perhaps be attributed to changes in national management guidelines to include

asymptomatic MSM with HIV and all individuals with LGV symptoms (Public Health England, 2020d). Between 2018 and 2019, there was a 21% increase in the number of LGV tests (from 10,502 to 12,674). There was also a corresponding increase in the percentage of LGV tests that were positive (from 8% to 9%), which indicated an increase in LGV transmission between 2018 and 2019.

In 2019, 98% (1,076/1,094) of all LGV diagnoses were in men, and diagnoses amongst women were rare (n= 18). Diagnoses in women may include transgender women. In men, LGV diagnoses in England are almost exclusively amongst MSM: 95% (1,020/1,076) of LGV diagnoses in 2019 were within this group. Men aged 25 to 44 accounted for 66% (n= 713) of all LGV diagnoses in 2019. Since 2018, fewer diagnoses are made in UK-born men compared to those born outside of the UK, and in 2019, 44% (478/1,076) of LGV diagnoses were made in UK-born men. The majority of LGV cases are white MSM (79% of all diagnoses in 2019), and a higher incidence of LGV is seen in MSM with a history of STI (Public Health England, 2020d). LGV diagnoses in the UK are concentrated in London (Public Health England, 2020d), with 77% of all LGV diagnoses in 2019 made in London. The rate of LGV is highest amongst MSM living with HIV however there has been an increase in LGV diagnoses amongst HIV-negative MSM since 2017. Between 2018 and 2019, there was a 95% increase in LGV diagnostic rate amongst HIV-negative MSM (Public Health England, 2020d). This group is eligible for pre-exposure prophylaxis (PrEP) (**1.8.3**). PrEP is the provision of antiretroviral drugs before HIV exposure to prevent infection. There are concerns that PrEP use may result in risk compensation, whereby individuals perceive themselves to be protected from HIV infection by PrEP, and as a result, become more likely to engage in riskier sexual practices that can increase the risk of STI transmission (McCormack *et al.*, 2016). Hence, the rise in LGV diagnoses amongst HIV-negative MSM may be the result of more widespread PrEP use in this population (Traeger *et al.*, 2018).

1.11 Overview of typing systems used to characterise *C. trachomatis* strains

Typing has been applied to differentiate *C. trachomatis* strains for decades. There are many uses and reasons for typing, including:

1. Differentiation of *C. trachomatis* strains can reveal transmission patterns in sexual networks and can be used as a tool in partner notification.
2. Typing may enable association with clinical manifestations and pathogenicity.
3. Tissue tropism for certain strains can be identified.
4. For evolutionary surveillance of specific clones.
5. Typing can play a role in cases of sexual abuse.
6. Typing may help to determine whether infections are persistent or new (Pedersen *et al.*, 2009)

In the 1970s, a microimmunofluorescence (MIF) test was developed to distinguish between *C. trachomatis* strains (Wang and Grayston, 1970). This test was able to differentiate strains into serovars (sometimes referred to as serotypes) based on variation in surface antigens. It was later recognised that these surface antigens were located on the chlamydial major outer membrane protein (MOMP) (**1.5**) (Stephens *et al.*, 1982). After the invention of PCR, serological classification of *C. trachomatis* was largely replaced by molecular typing methods based on the *ompA* gene encoding MOMP. One such method was restriction fragment length polymorphism (RFLP). This involved the amplification of the *ompA* gene using PCR, followed by DNA digestion using restriction endonucleases. The resulting DNA fragments were then separated on a polyacrylamide gel, and the pattern of DNA bands on the gel was compared against those of known serovars. It was later shown that sequencing of the *ompA* gene provided a higher level of resolution than serotyping using MIF, and RFLP; and the advent of molecular genotyping began.

The most widely used molecular genotyping methods for *C. trachomatis* are *ompA* genotyping based on PCR amplification and sequencing of the *ompA* gene; and high-resolution genotyping methods that target multiple loci across the *C. trachomatis* genome, including multi-locus sequencing typing (MLST) and multi-locus variable number tandem repeat (VNTR) analysis (MLVA). These techniques are described in detail in the following sections.

1.11.1 *OmpA* Genotyping

1.11.1.1 The *ompA* gene

The outer membrane protein A gene, *ompA*, encodes the *C. trachomatis* major outer membrane protein (MOMP). The *ompA* gene is approximately 1,200 base pairs in length and is comprised of five constant segments (CS 1 - 5) interspersed with four variable segments (VS 1 - 4). The constant segments encode five highly conserved membrane-spanning constant domains in the protein; and the variable segments encode four polymorphic variable domains (VD 1 – 4) that are surface-exposed epitopes (Stephens *et al.*, 1987). Amino acid substitutions within the variable domains of MOMP allows for *C. trachomatis* strain discrimination (**1.11**) (Spaargaren *et al.*, 2005).

1.11.1.2 Structure of the MOMP

The MOMP is embedded in the *C. trachomatis* outer membrane, and is considered to be a porin (Bavoil *et al.*, 1984). Porins are typically folded into a barrel-like β-pleated sheet structure delimiting a pore (Rodríguez-Marañón *et al.*, 2002). No experimental structural data exists for MOMP; however, models have proposed a trimeric structure consisting of 16 membrane-spanning β-sheets that form a barrel-like structure, with loops protruding outwards (Rodríguez-Marañón *et al.*, 2002). The variable domains are located on the outer loops, and the constant domains are located in the transmembrane region. The MOMP sequence has 8 – 10 cysteine residues that are located at the external membrane surface and in the external loops. Proposed functions for the cysteine residues include provision of stability to the outer *C. trachomatis* membrane through disulphide bond formation (Feher *et al.*, 2013). The predicted topology of MOMP is shown in Figure 6.

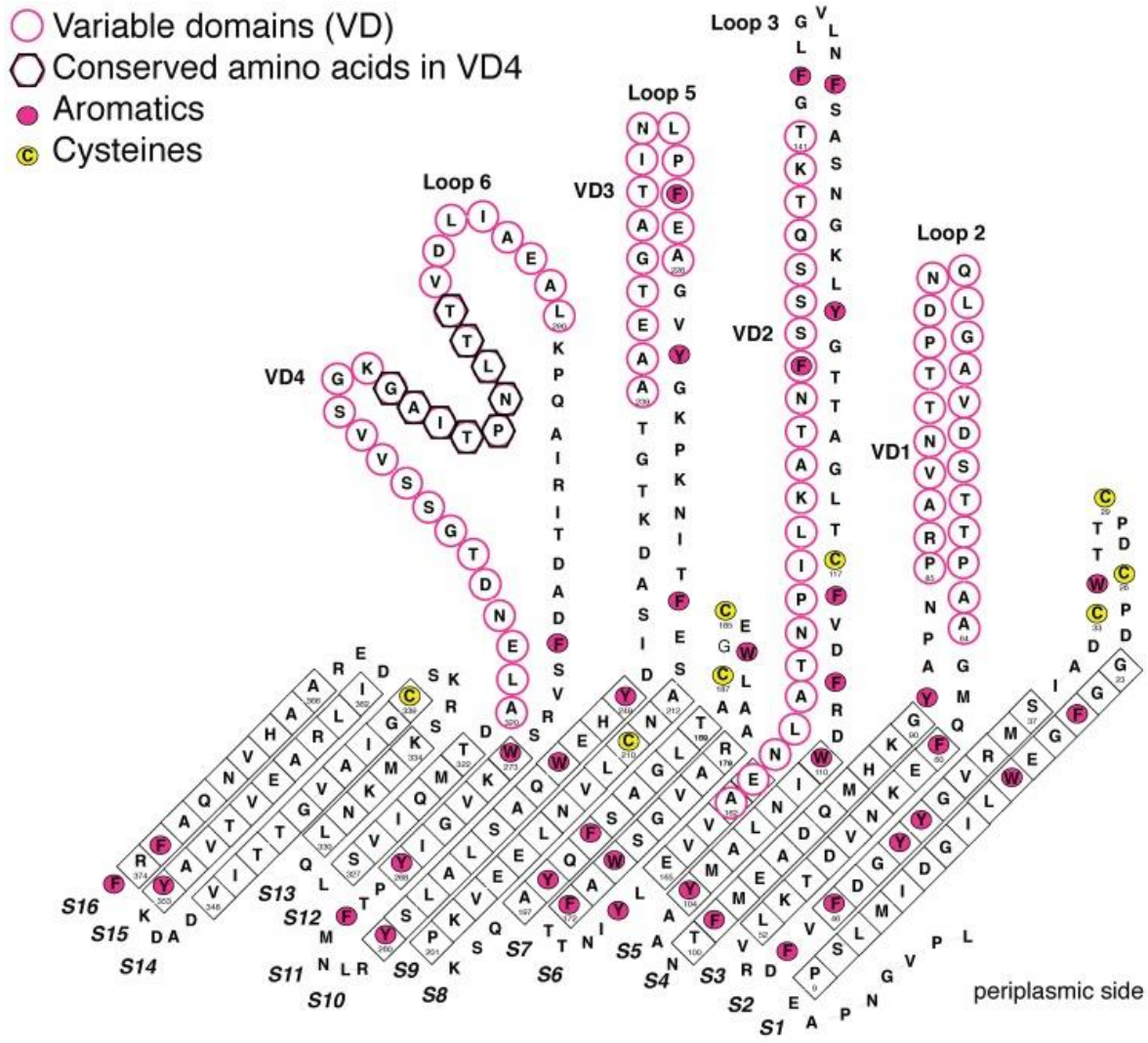


Figure 6 Predicted topology of the *C. trachomatis* major outer membrane protein (MOMP) in an *ompA* genotype C strain.

The major outer membrane protein (MOMP) of *C. trachomatis* consists of 16 anti-parallel beta-stranded sheets (S1 – S16), and six external loops. Within the loops, four surface-exposed variable domains (VD 1 – 4) are located. Five conserved domains (CD 1 – 5) are located in the transmembrane beta-sheet structure. The MOMP sequence has 8 – 10 cysteine residues (shown in yellow) that are located at the external membrane surface and in the external loops. Aromatic amino acid residues (shown in dark pink) are located in the periplasmic layer interface. Figure taken from (Feher *et al.*, 2013) and reproduced under the terms of the Creative Commons 4.0 licence.

1.11.1.3 *OmpA* genotyping

OmpA genotyping involves PCR amplification of the *ompA* gene using targeted primers, and sequencing of the resulting amplicons. Genotyping of the *ompA* gene is widely used to differentiate between *C. trachomatis* strains, however there are limitations to this method. *OmpA* genotyping lacks resolution and provides no information about the remainder of the *C. trachomatis* genome. Also, the *ompA* gene is not always an accurate epidemiological marker when used on its own, as *ompA* has been shown to be both a mutational and recombination hotspot within the genome (Brunelle and Sensabaugh, 2012; Harris *et al.*, 2012). For genotyping surveys, the method has been largely superseded by more discriminatory genotyping systems such as multi-locus sequence typing (MLST), and multi-locus variable number tandem repeat (VNTR) analysis (MLVA). Despite this, *ompA* genotyping is still the main method of *C. trachomatis* strain classification.

As detailed in **1.10.1**, sequence analysis of the *ompA* gene revealed that the LGV proctitis outbreak in MSM in Amsterdam was caused by a new *ompA* variant called L2b (Spaargaren *et al.*, 2005). The L2b variant has subsequently been implicated in outbreaks of LGV proctitis amongst MSM across the globe (Nieuwenhuis *et al.*, 2004; Kropp *et al.*, 2005; Ward *et al.*, 2007; Rodriguez-Dominguez *et al.*, 2015; Isaksson *et al.*, 2017). The new variant possessed several base pair changes to *ompA* genotype L2 and its known variants L2a and L2' (Spaargaren *et al.*, 2005). Three of these base pair changes encoded amino acid substitutions within variable domain 2 (VD 2) (**Figure 7**). It is unknown whether these substitutions cause functional alterations within the *ompA* gene. A fourth base pair change was identified in constant segment 2 (CS 2), however this mutation was synonymous and did not result in an amino acid substitution.

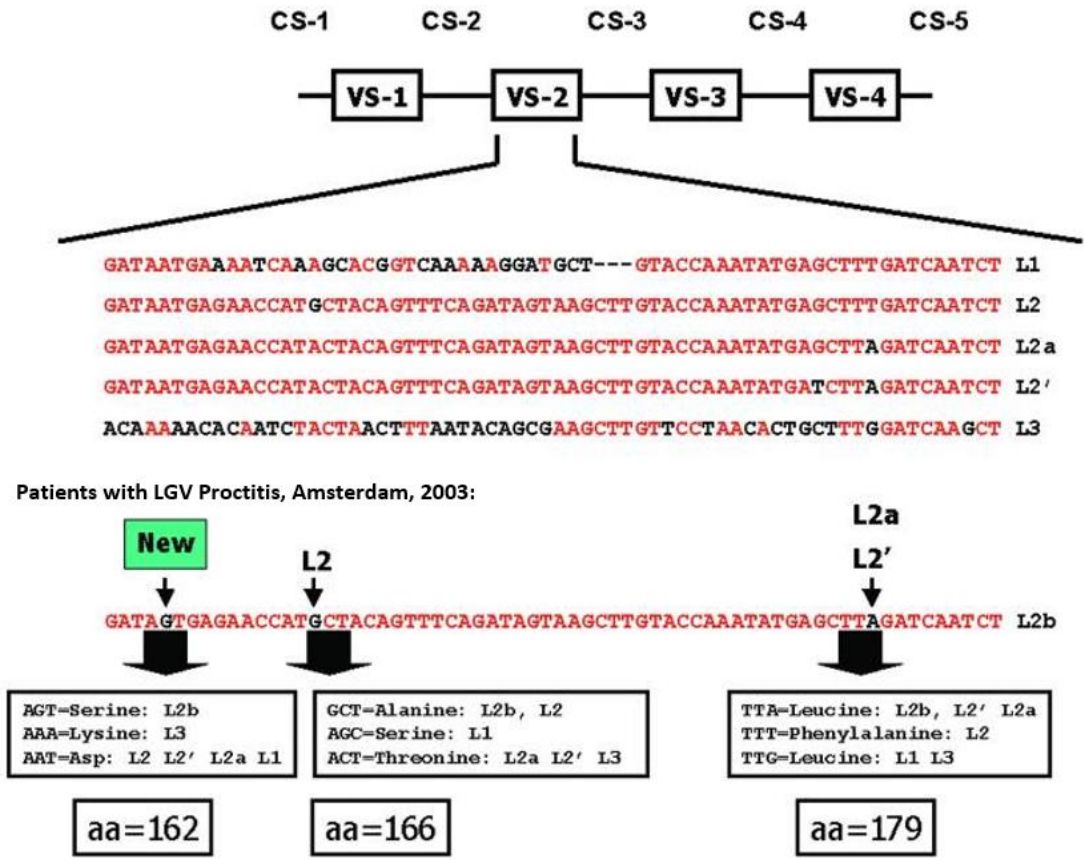


Figure 7 Schematic diagram of the *C. trachomatis ompA* gene.

For variable segment 2 (VS 2), the nucleotide and amino acid (aa) sequences of *ompA* genotypes L1, L2, L2a, L2' and L3 are compared to the L2b variant. Conserved nucleotides in VS 2 for all LGV strains are shown in red, and nucleotide substitutions are shown in black. The nucleotide substitutions in L2 as compared to L2, †L2a and †L2' are indicated by arrows. Amino acids encoded by the nucleotide substitutions, and the corresponding amino acid position are listed. Figure taken from Spaargaren *et al.* (2005) and reproduced under the Creative Commons 4.0 Licence. † L2a and L2' are *ompA* genotype L2 variants.

1.11.2 **Multi-locus Sequence Typing (MLST)**

Multi-Locus Sequence Typing (MLST) can be used to differentiate between *C. trachomatis* strains. MLST analyses nucleotide sequence variation within 400 - 700 base pair fragments of 6 - 8 housekeeping genes. Housekeeping genes are utilised as they are stable and are essential for cell survival, hence MLST can be used for long-term and global epidemiological studies. The genes chosen for MLST must not be under immune selection (Maiden, 2006), should be widely separated on the chromosome, and should not be adjacent to MOMP or hypothetical proteins as these may be under diversifying selection (Pannekoek *et al.*, 2008). MLST involves PCR amplification of the selected housekeeping genes followed by sequencing of the resulting amplicons. At each housekeeping gene locus, allele fragments are assigned a number, such that for every isolate, a code comprised of numbers for each locus is assembled. This code is referred to as a sequence type (ST). There are many advantages of MLST: its universal system of nomenclature allows inter-laboratory comparisons; and the use of nucleotide sequence analysis to determine STs is unambiguous and allows more alleles to be distinguished per locus, which affords a high level of isolate discrimination (Dean *et al.*, 2009).

In 2007, Klint *et al.* developed a multi-sequence typing scheme, termed MST, that targeted five non-housekeeping gene loci: CT046 (*hctB*), CT058, CT144, CT172 and CT682 (*pbpB*). The STs of 47 isolates were analysed and the scheme proved highly discriminatory, identifying 32 variants compared to the 12 variants detected using *ompA* sequence analysis (Klint *et al.*, 2007). This scheme does not follow the aforementioned rules of MLST - it is based on the sequencing of five hypervariable regions comprising genes that are under immune selection (*hctB* and *pbpB*; encoding DNA- and penicillin- binding proteins respectively). This scheme is often referred to as a multi-sequencing typing scheme, or MST (Labiran *et al.*, 2012).

In 2009, Pannekoek *et al.* developed an MLST scheme based on partial fragments (400 - 500 base pairs) of seven housekeeping genes (*gatA, oppA-3, hflX, gidA, enoA, hemN, fumC*), in order to investigate the population genetic structure of two chlamydial species, *C. trachomatis* and *C. pneumoniae* (Pannekoek *et al.*, 2008). The scheme provided no resolution for *C. pneumoniae* isolates, however ST analysis of 26 *C. trachomatis* isolates formed three clonal complexes consistent with those obtained from phylogenetic analyses of 16S and 23S rRNA genes (Everett, Bush and Andersen, 1999). LGV isolates formed one clonal complex; isolates with more common urogenital *ompA* genotypes (D - F) formed a second clonal complex; and isolates with less common urogenital *ompA* genotypes (G - K) and ocular *ompA* genotypes (A - C) formed the final clonal complex. These

findings were confirmed by analysis of whole genome sequence data (Harris *et al.*, 2012), and by the work of Dean *et al.* (2009). Dean *et al.* developed a second MLST scheme for *C. trachomatis*, and assigned STs to isolates from six continents using seven different housekeeping genes (*glyA, mdhC, pdhA, yhbG, pykF, lysS, leuS*). These isolates formed three clusters matching those reported by Pannekoek *et al.* (2008).

1.11.3 Multi-locus Variable Number Tandem Repeat (VNTR) Analysis (MLVA)

In 2008, Pedersen *et al.* detailed a new genotyping scheme for *C. trachomatis* isolate discrimination based on PCR and sequence analysis of the *ompA* gene, and three variable number tandem repeat (VNTR) regions (Pedersen *et al.*, 2008). This new genotyping scheme was termed multi-locus variable number tandem repeat (VNTR) analysis (MLVA) with *ompA* genotyping, or, MLVA-*ompA*. VNTRs are stretches of nucleotide repeats or motifs. In these regions of repetition, DNA polymerase is prone to error during replication. As such, VNTR targets are well-suited for short-term, local epidemiology studies. The three VNTR loci that comprise the Pedersen *et al.* (2008) scheme are CT1291, CT1299, and CT1335. These loci were selected on the basis that they had a high discriminatory power compared to other locus candidates (Pedersen *et al.*, 2008). The discriminatory power of a genotyping scheme is defined as its ability to assign a different genotype to two un-related strains sampled randomly from the population of a given species (Hunter and Gaston, 1988; van Belkum *et al.*, 2007). Discriminatory power is calculated using Simpson's Index of Diversity (**2.20.1**). The Index of Diversity can be used to compare typing schemes to allow for the selection of the most discriminatory system (Hunter and Gaston, 1988). A Simpson's index of diversity of 1.0 describes an ideal typing system, however a typing system with an Index value above 0.95 is considered to have more or less "ideal" resolution (van Belkum *et al.*, 2007). Pedersen *et al.* (2008) determined that for their study sample set, the use of *ompA* and the three VNTR loci gave a combined Simpson's index of diversity value of 0.94, which was higher than typing with just *ompA* (0.69) or each VNTR locus alone (0.53 - 0.74) (Pedersen *et al.,* 2008). Pedersen *et al.* (2008) also evaluated the typeability and reproducibility of the MLVA-*ompA* scheme. Reproducibility is the ability of a genotyping scheme to assign the same type to an isolate upon repeated testing. The reproducibility of MLVA-*ompA* was 100% (Pedersen *et al.*, 2008). Typeability is the ability of a typing system to assign a genotype to all isolates tested by the genotyping system. The typeability of *ompA* was 100% and 99% for MLVA-*ompA* (Pedersen *et al.*, 2008).

The VNTRs in the Pedersen *et al.* (2008) scheme are stable. Patients with recurrent *C. trachomatis* infections were shown to have unchanged VNTRs during the time of infection, a period ranging from 70 to 394 days (Pedersen *et al.*, 2008). This investigation had its limitations, namely that

patients were assessed under the assumption that each was infected with the same strain throughout the infection period, and also that each had not been re-infected with *C. trachomatis* by a different sexual partner. There were some concerns that the error-prone nature of DNA polymerase during DNA replication might affect the stability of the VNTR loci over several *C. trachomatis* divisions (Labiran *et al.*, 2012). This is crucially important as these loci must remain stable throughout an epidemiological study for the data being collected to be comparable (van Belkum *et al.*, 2007). The stability of the VNTR loci of the MLVA-*ompA* typing system was assessed by Labiran *et al.* (2012) by comparing the sequence at each VNTR locus prior to passage in tissue culture with the sequence at each locus after eight passages. It was found that the VNTR sequence after eight passages was identical to the VNTR sequence prior to passaging. This demonstrated that the MLVA-*ompA* typing system was suitable for both short- and long-term epidemiological investigations.

In 2012, Peuchant *et al.* developed a new MLVA scheme (MLVA-5) to discriminate between *C. trachomatis* isolates based on five VNTR loci (Ct-51, Ct-531, Ct-719, Ct-1025, and Ct-1035) (Peuchant *et al.*, 2012). MLVA-5 employs multiplex PCR and capillary electrophoresis with fluorescently-labelled primers and has no sequencing step. These methods enable high-throughput analysis and easier interpretation of results, particularly for VNTRs with a small number of repeat units (MLVA-5 repeat units vary from 3bp for Ct-719 and Ct-1025, to 150bp for Ct-531). The stability of MLVA-5 was assessed after 10 passages and the loci were shown to be stable (Peuchant *et al.*, 2012).

The ability of MLVA schemes to discriminate between *C. trachomatis* isolates makes them an attractive target for molecular epidemiological investigations of *C. trachomatis*.

1.12 Aims

As discussed in 1.11.1, *C. trachomatis* strains are classified by *ompA* genotype. The *ompA* gene has been identified as a mutational hotspot within the *C. trachomatis* genome, and extensive sequence variation within the *ompA* gene has been detected in *C. trachomatis* strains (Stary *et al.*, 2008; Brunelle and Sensabaugh, 2012; Borges *et al.*, 2019; Cole *et al.*, 2020b). Polymorphisms at other genetic loci within *C. trachomatis* also allows *C. trachomatis* strains to be differentiated. Polymorphisms within the tryptophan synthase alpha (*trpA*) gene have been found to distinguish trachoma-causing and genital *C. trachomatis* strains.

High-resolution genotyping systems targeting multiple loci across the *C. trachomatis* genome have increased our knowledge relating to genetic diversity within *C. trachomatis* strains. These multi-locus systems include MLST (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) and MLVA (Pedersen *et al.*, 2008; Peuchant *et al.*, 2012). A MLVA-*ompA* genotyping survey of *C. trachomatis* strains collected from the rectum of MSM in Brighton (Labiran *et al.*, 2016) demonstrated that a high level of genotypic diversity within *C. trachomatis* LGV strains exists within the Brighton population. However, the Brighton study included only eleven LGV isolates; and no large-scale high-resolution genotyping survey of LGV strains has ever been conducted in the UK. It is unknown whether this high level of genotypic diversity within LGV strains in Brighton is also present within LGV strains circulating in other UK cities.

Whole genome sequencing (WGS) has provided insight into the genomic diversity of *C. trachomatis* strains. Genomic analyses of LGV *C. trachomatis* strains have been conducted on strains isolated from patients with symptomatic LGV infections (i.e. the inguinal LGV syndrome or LGV proctitis). Strains isolated from patients with asymptomatic infections have not previously been analysed. It is unknown whether the absence of disease symptoms in asymptomatic individuals is the result of genetic variation within the genome, or a host effect. Further investigation into the genetic characteristics of asymptomatic LGV infections may reveal key marker sequences within the genomes.

This PhD will utilise several methods including *ompA* genotyping, *trpA* gene sequence analysis, MLVA-*ompA* genotyping, and whole genome sequencing; in order to characterise genetic diversity within *C. trachomatis* strains in high-risk sexual networks.

The aims of the PhD were as follows:

- To determine the *ompA* genotype of *C. trachomatis* DNA extracted from a conjunctival swab collected from a 10 year old female recently resident in Afghanistan, in order to

determine whether the child was infected with a trachoma-causing *C. trachomatis* strain, or a *C. trachomatis* strain associated with sexually transmitted urogenital infections (**Chapter 3**).

- To sequence and identify SNPs in the tryptophan synthase alpha subunit (*trpA)* gene of *C. trachomatis* DNA extracted from the conjunctival swab, to confirm that the child was infected with a trachoma-causing *C. trachomatis* strain (**Chapter 3**).
- To evaluate several multi-locus genotyping schemes for *C. trachomatis* in order to determine which scheme would be most suitable for a genotyping survey of LGV strains collected from London-based GUM clinics in Chapter 5 (**Chapter 4**).
- To consider the practicability and reproducibility of the most appropriate scheme for genotyping LGV strains of *C. trachomatis* (**Chapter 4**).
- To assess genotypic diversity within LGV strains in London, by determining the distribution of LGV MLVA-*ompA* genotypes in LGV-positive clinical specimens from eight London genitourinary medicine (GUM) clinics (**Chapter 5**).
- To compare the genomes of LGV isolates obtained from symptomatic and asymptomatic MSM in Brighton in order to investigate genomic diversity and indicate genomic loci within the isolates that may be linked with the presence or absence of clinical symptoms (**Chapter 6**).

CHAPTER 2 Materials and Methods

This chapter contains the standard methodologies of techniques used throughout this PhD. Specific details of experimental protocols and data analysis can be found in the corresponding methods section of the Results chapters.

2.1 *E. coli* growth media

2.1.1 Luria-Bertani (LB) Medium

LB medium is a nutritionally-rich medium used for the growth of *E. coli* strains (**Table 1**). Tryptone (10 g/L), Yeast Extract (5 g/L) and sodium chloride (5 g/L) (all *Becton, Dickinson and Company, USA*) were dissolved in UHQ dH_2O. Media was autoclaved. If required, ampicillin was added at a final concentration of 50 μg/ml once media had cooled (**2.1.4**). Bottles containing media were stored at room temperature.

2.1.2 LB Agar

LB agar plates were used for the growth and maintenance of *E. coli* strains (**Table 1**). Bacteriological Agar, 1.5% (*Becton, Dickinson and Company, USA*) was added to LB medium (**2.1.1**) and autoclaved. Molten LB agar was transferred to a water bath set to 55°C. If required, ampicillin was added at a final concentration of 50 μg/ml once media had cooled to 55°C (**2.1.4**). LB agar was poured into sterile petri dishes in the class II laminar flow cabinet. Once set, unused plates were inverted and stored at 4°C for a maximum of two weeks. Plates were visually checked for contamination prior to use in experiments.

2.1.3 Minimal Medium

Minimal medium was used to grow the thiamine-auxotrophic *E. coli* strains JM101 and DH5α (**Table 1**). The medium contained the minimum possible nutrients for colony growth supplemented with thiamine, to test thiamine dependency of the *E. coli* strains prior to their use in cloning. The purpose of this was to confirm the phenotype of each strain prior to use in cloning experiments.

Minimal E salts were dissolved in UHQ dH_2O overnight. The Minimal E salts were $MgSO_4.7H_2O$ (10 g/L), $C_6H_8O_7$ (100 g/L), K_2HPO_4 (500 g/L) and $NaNH_4HPO_4.4H_2O$ (175 g/L). Bacteriological agar, 1.5% (*Becton, Dickinson and Company, USA*) was dissolved in UHQ dH_2O ("water agar"), before autoclaving. Molten "water agar" was cooled to 55°C in a water bath. Per 100ml "water agar", 2 ml

of minimal E salts, 2 ml 10% glucose solution, and 200 µL 2.5 g/L thiamine hydrochloride were added. Minimal media was poured into sterile petri dishes in the class II laminar flow cabinet. Plates not supplemented with thiamine hydrochloride were also prepared. Once set, unused plates were inverted and stored at 4°C for a maximum of two weeks. Plates were visually checked for contamination prior to use in experiments.

2.1.4 Ampicillin

To grow the ampicillin-resistant *E. coli* strain JM101 (pUC19) (**Table 1**), media was supplemented with ampicillin. Ampicillin (*Merck, Germany*) was dissolved in UHQ dH_2O to 50 mg/ml. This was sterilised through a 0.2 µm syringe filter (*Thermo Fisher Scientific, UK*) and stored at -20°C in 500 µL aliquots. Ampicillin was added to cooled LB medium (**2.1.1**) or LB agar cooled to 55°C (**2.1.2**) at a final concentration of 50 µg/ml.

2.2 *E. coli* Strains

2.2.1 Growth and Storage of *E. coli* strains

A list of all *E. coli* strains used in this PhD is provided in **Table 1**. Frozen culture stocks of each *E. coli* strain were used to produce working stocks on agar plates supplemented with ampicillin (if required). Liquid cultures of *E. coli* were grown in LB media for 16 hours at 200rpm at 37°C. The culture was added to 30% sterile glycerol for long-term storage at -80°C. When required, a loop of bacteria was removed from the stock and streaked onto agar plates (supplemented with ampicillin (**2.1.4**) if required). For downstream applications, overnight cultures were prepared by inoculating 10 ml LB (unless otherwise stated) and growing for 16 hours at 37°C.

Table 1 ***E. coli*** **strains and features**

Strain	Genotype	Resistance	Reference
JM101	*supE thi-1 Δlac-proAB F'[lacI^{q}ZΔM15 traD36 proAB$^+$]*	None	(Messing, 1979)
DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80lacZΔM15 Δ(lacZYA-argF)U169 hsdR17(rk$^-$mk$^+$) Λ$^-$*	None	(Hanahan, 1983)
JM101 (pUC19)	As described for JM101, transformed with pUC19 vector	Ampicillin	(Yanisch-Perron, Vieira and Messing, 1985)

2.2.2 Preparation of competent *E. coli* cells

A 10 ml culture of *E. coli* was grown overnight at 37°C, 5% CO_2 (**2.2.1**). A 1 ml aliquot of overnight culture was used to inoculate 100 ml of LB media (supplemented with ampicillin if required; **2.1.4**). This was incubated at 37°C, 5% CO_2, 200rpm to log phase (OD_{600nm} = 0.3). The culture was placed on ice for 20 minutes. Cells were pelleted by centrifugation at 4,000 ×g for 10 minutes. Supernatant was discarded and cells re-suspended on ice in 20 ml of ice-cold 0.1M $CaCl_2$ for 20 minutes. Cells were re-pelleted by centrifugation at 4,000 ×g for 10 minutes. Supernatant was discarded and cells re-suspended in 2ml of ice-cold 0.1M $CaCl_2$ with 20% glycerol. Competent cells were stored at -80°C in 200 µL aliquots. Cells were thawed on ice prior to transformation.

2.3 Plasmids

The plasmid vector, pUC19, was used for cloning in this PhD (**Table 2**). This vector is 2,686 base pairs (bp) in length and has a high (500-700) copy number. A vector map for pUC19 is provided in Figure 8. Purified pUC19 stocks were prepared using the Wizard® Plus SV Minipreps DNA Purification System (*Promega, UK)* (**2.4.1**) and stored at -20°C.

Table 2 Plasmids

Plasmid	Size (bp)	Features	Resistance	Reference
pUC19	2,686	Cloning vector with a 54 base pair multiple cloning site (MCS), an Ampicillin resistance gene (*Amp*^{R}), and the N-terminus fragment of the β-galactosidase-encoding gene *lacZ (lacZα).*	Ampicillin	(Norrander, Kempe and Messing, 1983)

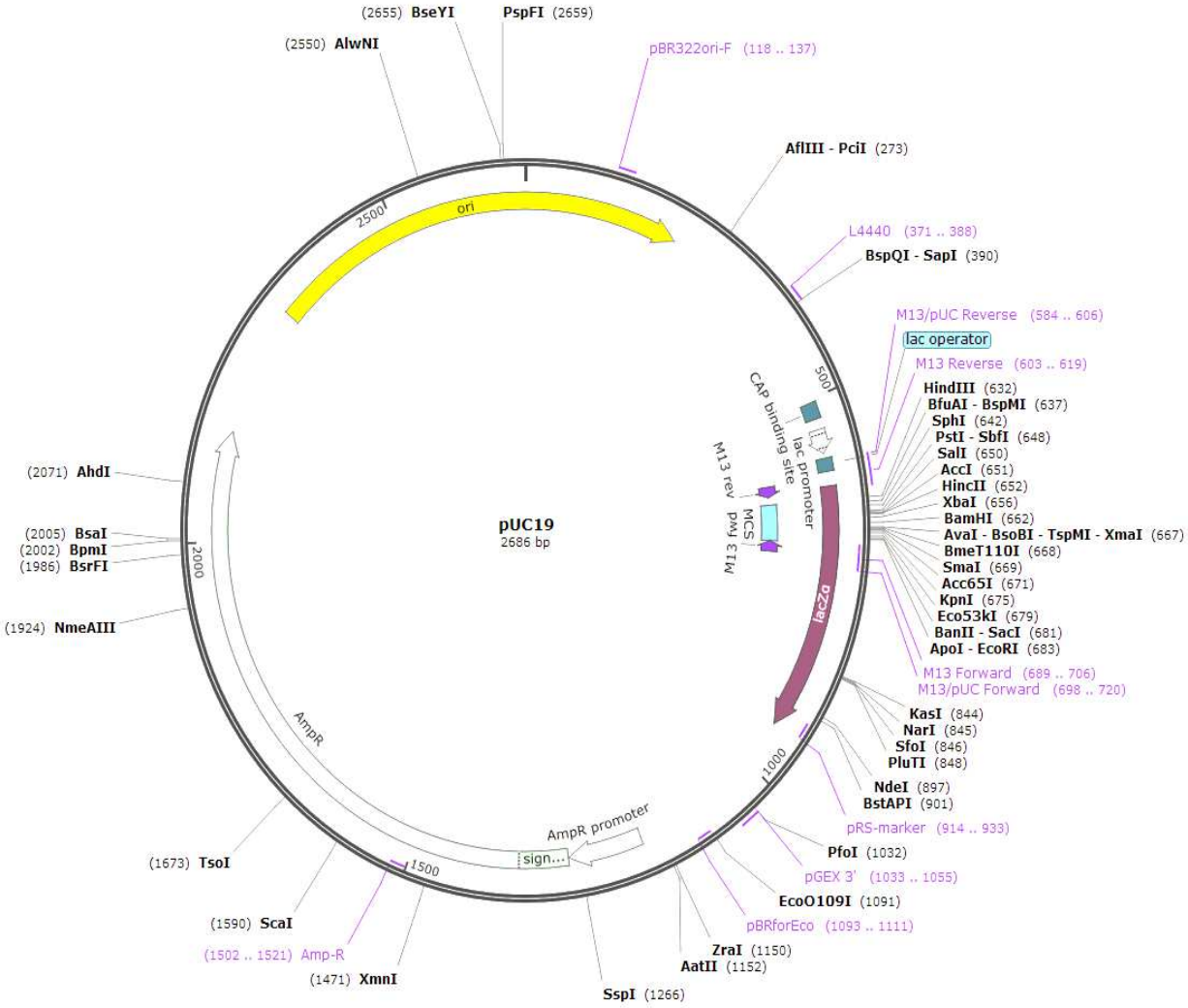


Figure 8 pUC19 Vector Map.

The multiple cloning site (MCS), the Ampicillin resistance gene (AmpR), the origin of replication (ori), the CAP binding site and lac promoter, and the lacZα gene are shown, along with M13 forward and reverse sequencing primer binding sites and restriction sites. Vector map was created by Addgene (Addgene plasmid #50005) using SnapGene® (version 5.3).

2.4 Isolation of DNA

2.4.1 Isolation of Plasmid DNA

Plasmid DNA was isolated from *E. coli* using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*). Bacteria was inoculated into 4 ml of LB medium (with 50 µg/ml ampicillin) and incubated overnight at 200rpm at 37°C (**2.2.1**), and pelleted by centrifugation at 10,000 ×g for 5 minutes. Supernatant was discarded and the pellet re-suspended in 250 µL of Cell Resuspension Solution. Cell Lysis Solution (250 µL) was added and the mixture inverted 4 times to mix. Alkaline Protease Solution (10 µL) was added to inactivate endonucleases released during the cell lysis step that degrade DNA. The mixture was then inverted a further 4 times before incubation at room temperature for 5 minutes. Neutralisation solution (350 µL) was added and the mixture inverted 4 times to mix, before centrifugation at 16,000 ×g for 10 minutes. The lysate was decanted into a Spin Column sat atop a Collection Tube, and centrifuged at 16,000 ×g for 1 minute. The flow-through was discarded. The silica membrane of the Spin Column was washed twice using Wash Solution. For the first wash, 750 µL of Wash Solution was added to the membrane and the Spin Column centrifuged at 16,000 ×g for 1 minute. For the second wash, 250 µL of Wash Solution was added and the Spin Column centrifuged at 16,000 ×g for 2 minutes. After both wash steps the flow-through was discarded. The Spin Column was transferred to a sterile 1.5 ml microcentrifuge tube. The final step was to elute the DNA in 100 µL of Nuclease-Free Water (*Life Technologies, UK*) and centrifuge at 16,000 ×g for 1 minute.

Plasmid DNA concentrations were determined by dsDNA fluorometry using Qubit® (*Thermo Fisher Scientific, UK)* (**2.12**) to inform the required volume for downstream applications. Isolated plasmid DNA was analysed by agarose gel electrophoresis (**2.11**). If not immediately required for downstream applications, DNA was stored at -20°C.

2.4.2 Isolation of Genomic DNA

*G*enomic DNA was isolated from *C. trachomatis* inocula (**2.17**), and from clinical specimens including swabs, using the NucleoSpin® Tissue kit (*Macherey-Nagel, Germany*).

Lysis buffer (T1; 180 µL) and 25 µL of Proteinase K were added to clinical samples before vortexing. This mixture was incubated in a water bath set to 56°C for 1 - 3 hours. An RNase digest was performed by adding 20 µL RNase A (10 mg/ml) and incubating at room temperature for 5 minutes. Samples were vortexed and 200 µL of lysis buffer (B3) was added, before vortexing and incubating at 70°C for 10 minutes. Ethanol (100%; 210 µL) was added before vortexing and transferring the resulting solution to a NucleoSpin® Tissue Column. The tissue column was centrifuged at 11,000 ×g

for 1 minute and the flow-through discarded. The silica membrane of the Tissue Column was washed twice using wash buffers. For the first wash, 500 µL buffer BW was added to the membrane and the column centrifuged at 11,000 ×g for 1 minute. For the second wash, 600 µL buffer B5 was added and the column centrifuged at 11,000 ×g for 1 minute. After both wash steps the flow-through was discarded. The silica membrane of the Column was dried to remove residual ethanol by centrifuging the column at 11,000 ×g for 1 minute. The Tissue Column was transferred to a sterile 1.5 ml microcentrifuge tube. DNA was eluted by addition of 100 µL of elution buffer (BE) containing 5 mM Tris-HCl, and incubating at room temperature for 1 minute, before centrifugation at 11,000×g for 1 minute.

Genomic DNA concentrations were determined by dsDNA fluorometry using Qubit® (*Thermo Fisher Scientific, UK)* (**2.12**) to inform the required volume for downstream applications. If not immediately required for downstream applications, DNA was stored at -20°C.

2.5 Primers

Primers (*Eurogentec, Belgium*) were used to amplify sections of plasmid or genomic DNA by PCR. For MLVA-*ompA* genotyping (**Chapter 4** and **Chapter 5**), primers designed by Pedersen *et al.* (2008), Labiran *et al.* (2014) and Lan *et al.* (1994) were used. For cloning, custom primers were designed using SnapGene® (version 5.3) with restriction sites (**2.6**) to facilitate ligation (**2.7**) of amplified DNA products into plasmid vectors.

2.6 DNA Digestion

Inserts generated by PCR (**2.9**), and plasmids (**2.3**), were digested at restriction sites using restriction endonucleases. If required, restriction sites were added to the termini of the insert DNA using PCR with cloning oligonucleotides (**2.5**). DNA (2 µg) was incubated for 2 hours at 37°C with 1x digest buffer (*Promega, UK)*, 5 units of the relevant endonuclease (*Promega, UK)*, and 1x BSA (*Promega, UK*).

Linearised plasmid vectors were dephosphorylated prior to cloning to prevent re-circularisation during ligation. The dephosphorylation reaction contained 5 units of alkaline phosphatase (*Promega, UK*), 1x reaction buffer (*Promega, UK*), and 1x BSA (*Promega, UK*). This reaction was incubated at 37°C for 30 minutes then 56°C for 15 seconds for deactivation. Digested vector DNA was purified (**2.4.1**) prior to ligation (**2.7**).

2.7 Ligation

Ligations were performed using T4 DNA Ligase (*Promega, UK*). T4 DNA Ligase facilitates ligation by catalysing the formation of a phosphodiester bond between the 5' phosphate and the 3' hydroxyl termini of two separate DNA strands. Each 10 µL ligation reaction consisted of 10 ng dephosphorylated vector (**2.6**), a 3-fold molar excess of insert DNA, 1 unit of T4 DNA Ligase and 1x Rapid Ligation Buffer (*Promega, UK*). Reactions were incubated at room temperature for 15 minutes. Ligated plasmids were transformed (**2.8**) into competent *E. coli* cells.

2.8 Transformation of *E. coli*

Ice-cold plasmid vector DNA (10 ng) (**2.7**) was added to 50 µL of competent *E. coli* cells (**2.2.2**) and kept on ice for 25 minutes. Cells were heat-shocked at 42°C for 45 seconds to facilitate the uptake of DNA, then placed on ice for 5 minutes to recover. LB medium (500 µL; **2.1.1**) was added to the cells and the mixture incubated at 37°C, 5% CO_2, 200rpm for 1 hour. Cells were pelleted by centrifugation at 3,000 ×g for 1 minute at 20°C. The supernatant was discarded and the cells re-suspended in 50 µL of LB medium. Transformants were screened (**2.8.1**) on LB agar supplemented with 5-bromo-4-chloro-3-indoyl-beta-D-galacto-pyranoside (X-Gal), isopropyl beta-D-1-thiogalactopyranoside (IPTG) and ampicillin (**2.1.4**) and incubated overnight at 37°C.

2.8.1 Screening for Transformants

Transformed *E. coli* colonies were identified by blue-white screening. A schematic diagram of the blue-white screening process is provided in Figure 9. Briefly, *E. coli* strains with the lacZΔM15 deletion mutation are unable to produce functional β-galactosidase. Upon transformation with a plasmid containing a segment of *lacZ* (*lacZα*), functional β-galactosidase is produced. This process is called α-complementation. A multiple cloning site (MCS) is present within the *lacZ* gene in the plasmid. When *E. coli* is transformed with a recombinant plasmid containing foreign DNA within the MCS, α-complementation cannot occur and functional β-galactosidase is not produced.

For screening of transformed *E. coli*, LB agar (**2.1.2**) was supplemented with X-Gal (20 µg/ml), IPTG, and Ampicillin (50 µg/ml). IPTG induces the expression of *lacZ.* If functional β-galactosidase is produced, X-Gal is hydrolysed to form 5-bromo-4-chloro-indoxyl, which spontaneously dimerises to produce an insoluble blue pigment called 5-5'-dibromo-4,4'-dichloro-indigo. Colonies transformed with non-recombinant plasmid appeared blue whilst those transformed with recombinant plasmid (insert within MCS) appeared white. This process is shown in Figure 9.

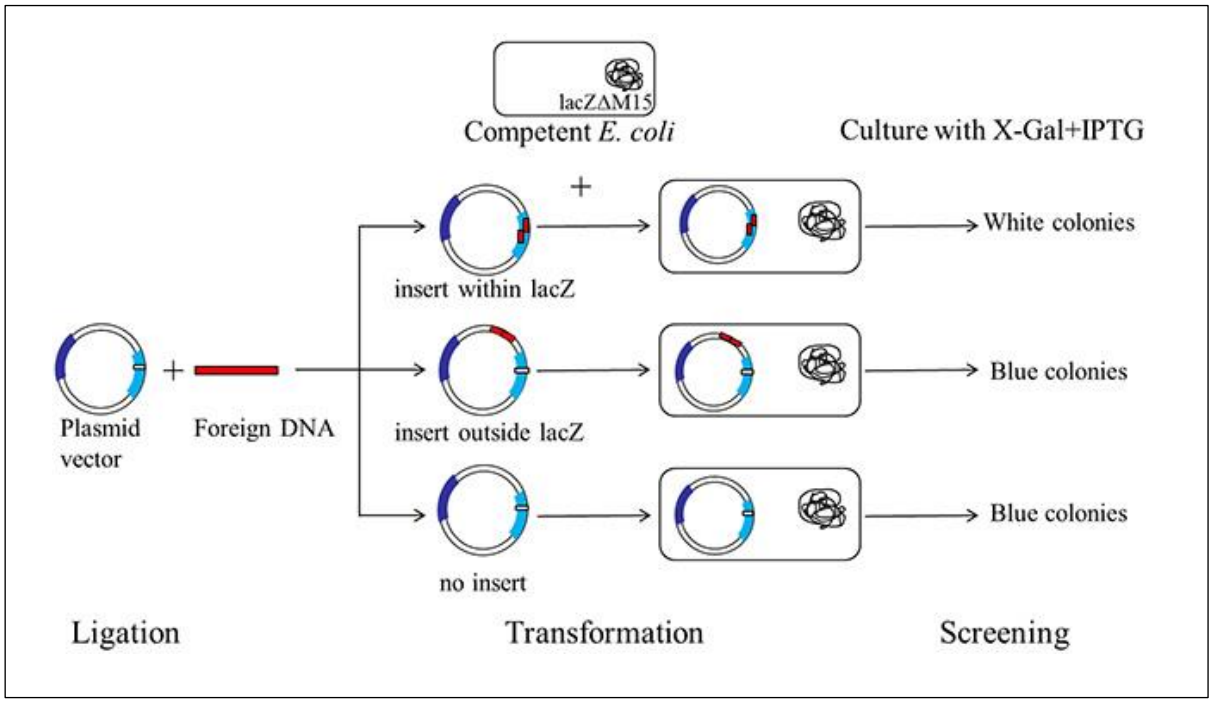


Figure 9 Blue-white screening of *E. coli* transformants.

Plasmid vector (with *ampR* in dark blue, lacZα in light blue, and the multiple cloning site in white within lacZα) is ligated with foreign DNA and transformed into competent *E. coli* cells with a lacZΔM15 deletion mutation. *E. coli* is grown on agar plates supplemented with X-Gal and IPTG (and antibiotics if required). The plasmid vector has a *lacZα* gene that encodes a functional β-galactosidase enzyme when transformed into *E. coli* strains with a lacZΔM15 deletion mutation. A multiple cloning site (MCS) is present within the *lacZα* gene in the plasmid. Functional β-galactosidase is not produced when foreign DNA is inserted at the MCS within *lacZα*. β-galactosidase expression is induced by IPTG. β-galactosidase hydrolyses X-Gal to form an insoluble blue pigment. *E. coli* cells transformed with plasmid vector containing either no foreign DNA, or foreign DNA inserted outside of *lacZα* produce functional β-galactosidase and colonies are blue. *E. coli* cells transformed with plasmid vector containing foreign DNA inserted within *lacZα* do not produce functional β-galactosidase and colonies are white. Figure from (Merck, 2020).

2.9 DNA Amplification

Template DNA was amplified by Polymerase Chain Reaction (PCR). PCR amplification was conducted using a Veriti™ 96-well thermal cycler (*Applied Biosystems, UK).*

2.9.1 Amplification of VNTR and *ompA* sequences

OmpA was amplified using primers PCTM3 and NRI (Lan *et al.*, 1994) (**Table 3**), whilst the VNTR regions (CT1335, CT1299, and CT1291) were amplified using the primers detailed in Pedersen *et al.* (2008): CT1335F and CT1335R, CT1299F and CT1299R, and CT1291F and CT1291R (**Table 3**). Extracts that did not produce VNTR amplicons using these primers were amplified using primers designed by Labiran (2014): CT1335F* and CT1335R*, CT1299F* and CT1299R*, CT1291F* and CT1291R* (**Table 3**). The forward primers (marked with F*) annealed upstream of the original VNTR amplicon sequences, and the reverse primers (marked with R*) downstream, so that the original amplicon sequences were encompassed by the primers.

Each PCR was conducted in 20 µL reactions consisting of: 10 µL Phusion Flash High-Fidelity PCR Master-Mix (*Thermo Fisher Scientific, UK)*, 0.5 µM of forward and reverse primers (*Eurogentec, Belgium)*, 1 µL DNA, and UltraPure™ DNase/RNase-Free distilled H_2O *(Life Technologies, UK)*. Phusion Flash is a 2X master mix that utilises a high-fidelity proof-reading DNA polymerase called Phusion Flash II DNA Polymerase, which ensures a low error-rate.

The PCR conditions used to amplify VNTR and *ompA* sequences are provided in Table 4. As described in Pedersen *et al.* (2008) and Wang *et al.* (2011), 35 cycles of denaturation, annealing and extension was carried out for *ompA;* and 40 cycles for each VNTR region.

Table 3 Primer sequences for PCR amplification of VNTR loci and *ompA*

Primer Name	Primer sequence (5'- 3')	Amplicon size[a] (bp)	Nucleotide position[a]	Reference
CT1335F	TCATAAAAGTTAAATGAAGAGGGACT	153	737,225–737,250	(Pedersen *et al.*, 2008)
CT1335R	TAATCTTGGCTGGGGATTCA		737,377–737,358	
CT1299F	TTGTGTAAAGAGGGTCTATCTCCA	188	291,758–291,781	
CT1299R	AAGTCCACGTTGTCATTGTACG		291,945–291,924	
CT1291F	GCCAAGAAAAACATGCTGGT	225	195,536–195,555	
CT1291R	AGGATATTTCCCTCAGTTATTCG		195,760–195,738	
PCTM3	TCCTTGCAAGCTCTGCCTGTGGGGAATCCT	1,019	779,977–780,006	(Lan *et al.*, 1994)
NR1	CCGCAAGATTTTCTAGATTTC		779,008–778,988	
CT1335F*	AAAGCGTCCTCTGGAAGGG	208	737,198-737,216	(Labiran, 2014)
CT1335R*	CCTTCTCCTAACAACTTACGC		737,405-737,385	
CT1299F*	ATCGCTTAAGATTCTCGGAGG	342	291,654-291,674	
CT1299R*	AGGTTCTAGCTGAGCATGGG		291,995-291,976	
CT1291F*	ATATAAAAAGAACCGTTGTTTCTG	329	195,462-195,485	
CT1291R*	CATCTTAGACATGCTCCGGC		195,790-195,771	

[a] According to D/UW-3/CX, Genbank accession number NC_000117

Table 4 PCR conditions for amplification of VNTRs and *ompA*

Denature		98°C	10 secs
	Repeat 35 - 40 times*	98°C	2 secs
Anneal		55°C (CT1335) 64°C (CT1299) 59°C (CT1291 and *ompA*)	5 secs
Extend		72°C	10 secs
		72°C	1 min
		10°C	∞

*Thermal Cycling: 35 cycles for *ompA*; 40 cycles for VNTR regions.

2.10 DNA Purification

PCR products (**2.9**) were purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega, UK)* prior to sequencing (**2.13**). DNA bands were excised from agarose gels (**2.11**) using Swann-Morton® sterile surgical scalpel blades *(Swann-Morton®, UK)* and weighed. Membrane Binding Solution was added at 10 µL per 10 mg of gel slice, and incubated at 56°C until the gel slice had completely dissolved (approximately 10 minutes). Dissolved gel mixtures were vortexed before adding to a Minicolumn and incubating at room temperature for 1 minute. The Minicolumn was centrifuged at 16,000 ×g for 1 minute and the flow-through discarded. Membrane Wash Solution (containing 95% ethanol; 700 µL) was added and the Minicolumn centrifuged at 16,000 ×g for 1 minute before discarding the flow-through. Membrane Wash Solution (500 µL) was added and the Minicolumn assembly centrifuged at 16,000×g for 2 minutes before discarding the flow-through. To evaporate any residual ethanol, the Minicolumn assembly was centrifuged at 16,000 ×g for a further two minutes. The Minicolumn was transferred to a 1.5 ml microcentrifuge tube and 50 µL UltraPure™ DNase/RNase-Free distilled H_2O *(Life Technologies, UK)* was added and incubated at room temperature for 1 minute before centrifugation at 16,000 ×g for 2 minutes.

Purified DNA concentrations were determined by dsDNA fluorometry using Qubit® (*Thermo Fisher Scientific, UK)* (**2.12**) to inform the required volume for downstream applications. If not immediately required for downstream applications, purified PCR products were stored at -20°C.

2.11 Agarose Gel Electrophoresis

2.11.1 Materials

Tris-acetate EDTA buffer (TAE; 50x) (*Thermo Fisher Scientific, UK*) prepared at a 1x working dilution was used as the electrophoresis buffer for agarose gels. DNA Loading Buffer Blue (5x; *Bioline, UK*) was used as the sample loading buffer to track the migration of DNA through the agarose gel. SYBR™ Safe DNA Gel Stain (10,000x; *Invitrogen, UK)* was added to molten agarose gels (**2.11.2**) to visualise DNA in agarose gels under UV light. SYBR™ Safe was selected as it was the safer alternative to ethidium bromide for DNA visualisation. DNA ladders including HyperLadder™ I and II (*Bioline, UK*) and SmartLadder (*Eurogentec, Belgium*) were used during the PhD as markers to estimate the size (in bp) of separated DNA fragments on agarose gels (**Figure 10**) (**2.11.2**).

2.11.2 Agarose Gel Electrophoresis

Agarose (*Thermo Fisher Scientific, UK*) gel electrophoresis was used for the separation of DNA fragments. Agarose (1 - 2%) was dissolved in 1x TAE buffer by heating to 100°C in a microwave oven. SYBR™ Safe DNA Gel Stain was added to a final concentration of 1x. The molten agarose was poured into a gel mould with a comb and allowed to set at room temperature for 30 minutes prior to use. The comb was removed and the gel was submerged in 1x TAE buffer. DNA samples containing 1x DNA Loading Buffer Blue were loaded into wells. DNA ladder (5 µL) was loaded onto the gel. Gels were electrophoresed at 90 volts until the blue dye line (the DNA Loading Buffer Blue) had migrated to 1cm from the end of the gel. Gels were imaged under UV light using the Gel Doc™ XR+ Transilluminator (*Bio-Rad, UK*). The sizes of separated DNA fragments were estimated by comparison with bands of known size using the DNA ladders Hyperladder™ I (200 – 10,000 bp) (*BioLine, UK*), HyperLadder™ II (50 – 2,000 bp) (*BioLine, UK*) or SmartLadder (200 – 10,000 bp) (*Eurogentec, Belgium*) (**Figure 10**). Hyperladder™ I and SmartLadder were used interchangeably depending on availability in the laboratory.

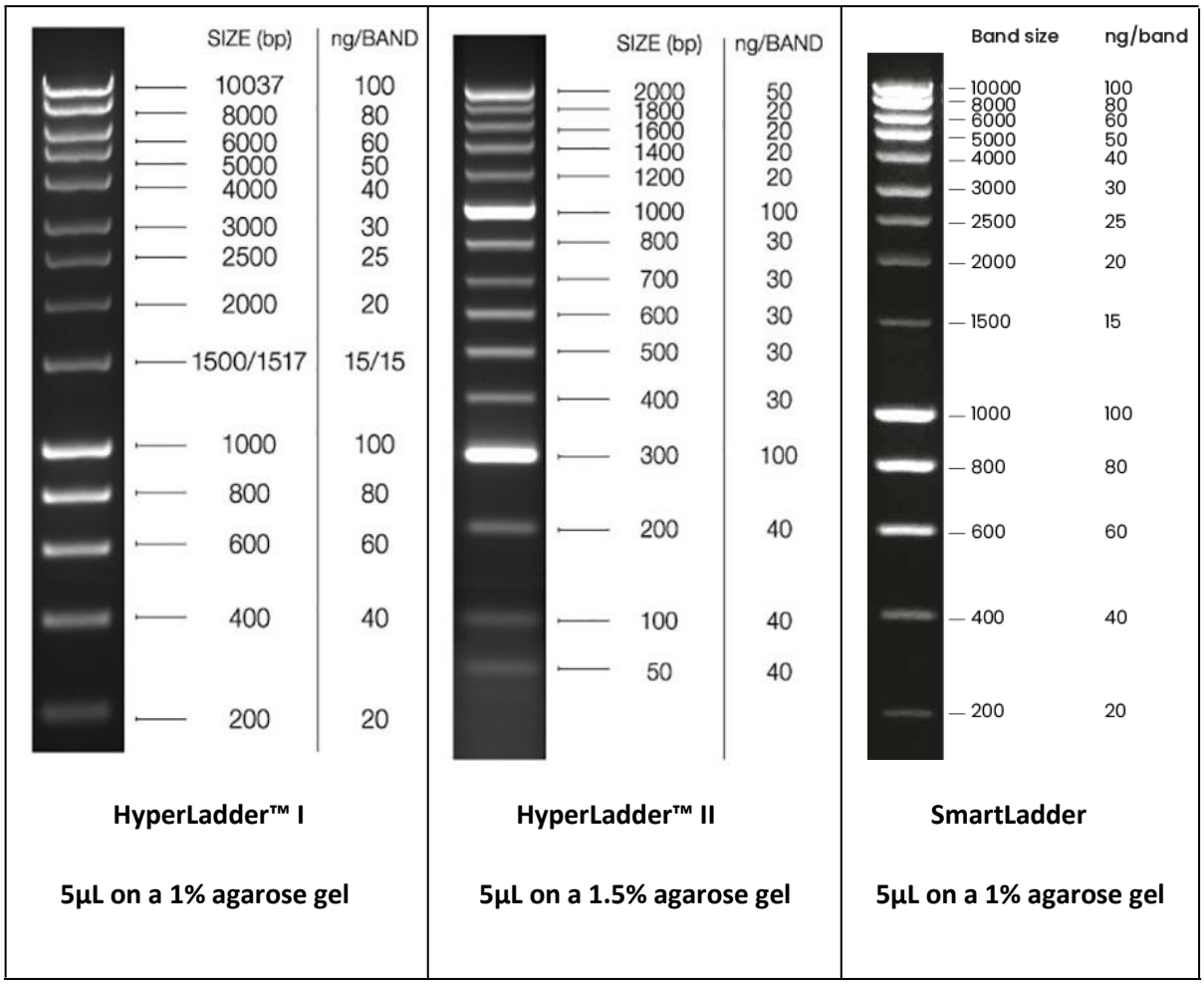


Figure 10 DNA Ladders used to estimate size (bp) of separated DNA fragments in agarose gels.

HyperLadder™ I (Left), HyperLadder™ II (Centre), and SmartLadder (Right). Each DNA ladder has several spaced bands ranging in size from 200 - 10,000 bp (HyperLadder™ I and SmartLadder) or 50 - 2,000bp (HyperLadder™ II). Each spaced band corresponds to an exact quantity of DNA (in ng/band) from 15 – 100 ng/band (HyperLadder™ I and SmartLadder) or 20 – 100 ng/band (HyperLadder™ II).

2.12 DNA Quantitation using fluorometry

DNA concentrations were determined by fluorometry using the Qubit® dsDNA high-sensitivity (HS) assay (*Invitrogen, UK)* with a Qubit® 1.0 Fluorometer (*Invitrogen, UK)*. This assay utilises dyes that emit fluorescence when bound to a specific target: double-stranded DNA (dsDNA). The fluorometer records the signal intensity and compares it to two quantitative standards to achieve a read-out of sample DNA concentration. The dsDNA assay is highly sensitive, as the fluorescent dyes do not bind to other molecules such as RNA or contaminants including salts and proteins.

2.12.1 **Qubit® dsDNA High-Sensitivity (HS) Assay**

The Qubit® dsDNA high-sensitivity (HS) assay is accurate for sample concentrations ranging from 10 pg/µL to 100 ng/µL.

A Qubit® working solution was prepared by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer. The required number of tubes for standards and samples were labelled. The Qubit® dsDNA HS assay uses 2 standards. The final volume in each tube was 200 µL. To the tubes labelled "Standard 1" and "Standard 2", 190 µL of Qubit® working solution was added, in addition to 10 µL of each Qubit® standard to the appropriate tube. To each sample tube, 199 µL of Qubit® working solution was added, followed by 1 µL of sample. Each tube was mixed by vortexing for 2 - 3 seconds. Tubes were left to incubate at room temperature for 2 minutes.

On the Home screen of the Qubit® 1.0 Fluorometer, "dsDNA HS assay" was selected followed by "Run New Calibration". Standard 1 was inserted and "Go" pressed. Standard 1 was replaced with Standard 2, and "Go" pressed. After the calibration was complete, the first sample was inserted, and "Go" pressed, followed by "Calculate sample concentration" to record the concentration.

2.13 **Sanger Sequencing**

Purified PCR products (**2.10**) and recombinant plasmid vectors (**2.8**) were sequenced by Source Bioscience Cambridge. Sample requirements were as follows: 10 ng/µL for PCR products; 100 ng/µL for recombinant plasmids; and 3.2 pmol/µL for primers. All samples sent for sequencing were purified using purification kits (**2.4.1**for recombinant plasmids and **2.10** for PCR products). Recombinant plasmids were sequenced using M13 forward and reverse sequencing primers provided by Source Bioscience. VNTR and *ompA* PCR products were sequenced using the primers used for PCR amplification (**Table 3**).

2.14 McCoy Cell Culture

2.14.1 McCoy Cell line

C. trachomatis cell culture was carried out using the mouse fibroblast cell line, McCoy B. McCoy B cells were obtained from the European Collection of Authenticated Cell Cultures (ECACC, catalogue number 90010305, *Public Health England, UK*). Cells were routinely cultured in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L glucose, L-glutamine, and no sodium pyruvate (*Life Technologies, UK*), supplemented with 10% foetal calf serum (FCS; *Life Technologies, UK*). Cells cultured in a T75 flask were maintained in a 15ml volume of media and were passaged twice weekly.

2.14.2 Cell Passage

Monolayers of McCoy cells were passaged when sub-confluent (i.e. 80 - 90% of the surface area of the bottom of the flask was covered by a layer of cells). Sub-confluence was achieved approximately 24 - 48 hours after incubation at 37°C in 5% CO_2 with DMEM supplemented with 10% FCS. Prior to each passage, the necks of tissue culture flasks and bottles were wiped with 70% ethanol to prevent contamination. Spent media was removed and cells washed twice in phosphate buffer saline (PBS; *Life Technologies, UK*) to remove traces of FCS in the media that inhibit trypsinisation. Flasks were incubated for 5 minutes with 1.5 ml 0.25% Trypsin/EDTA (TE; *Life Technologies, UK*) in a 37°C incubator with 5% CO_2 to facilitate cell detachment. Fresh media (volume dependent on split ratio desired) was added to inhibit trypsin, and cells were washed up and down the flask with the media by pipetting several times to break up clumps of cells. New T75 flasks were labelled with the date and split ratio, and 15 ml of DMEM (supplemented with 10% FCS) was added to each. A calculated volume of the DMEM-cell mixture was added to the new flask. The caps of the T75 flasks were loosened to allow gas exchange, and flasks were incubated in a 37°C incubator with 5% CO_2.

2.15 *Chlamydia* Species

All work was conducted using the *Chlamydia* species *C. trachomatis*.

2.16 Infection of McCoy cell monolayer with *C. trachomatis*

McCoy cells were seeded into a T25 flask 24 hours prior to infection with *C. trachomatis.* Cells were grown overnight in DMEM supplemented with 10% FCS and incubated in a 37°C incubator with 5% CO_2. The following day, flasks were removed from the incubator and spent media discarded. Fresh media (5 ml) containing bacterial inocula, cycloheximide (1 µg/ml), gentamycin (20 µg/ml), and

vancomycin (10 µg/ml) (all *Sigma-Aldrich, UK*) was added to the flask. For LGV *C. trachomatis* strains, cells were incubated in a 37°C incubator with 5% CO_2 for 48 hours with the flask cap loosened to allow for gas exchange. Cells were routinely visualised by phase contrast microscopy using a Nikon eclipse TS100 inverted microscope with 10x, 20x, and 40x objectives and fluorescence accessories. Images were captured using a Nikon DS-Fi1 camera head.

2.17 Harvesting of *C. trachomatis*

Once mature inclusions were visible in the cell monolayer (~36 - 48 hours after infection with *C. trachomatis*), *C. trachomatis* was harvested from the tissue culture flasks. A cell scraper was used to detach the cells from the plastic of the flask into the medium. The media was subsequently decanted into a 50 ml universal. The universal was centrifuged at 2850 ×g for 10 minutes to separate the cells (in pellet) from the media (in supernatant). The supernatant was discarded and the pellet re-suspended in 2 ml 10% ice-cold PBS to lyse the cells. The re-suspended pellet was transferred to a bijoux containing ~5 glass beads and vortexed at high speed for 1 minute to release *C. trachomatis* EBs from the cells. The contents of the bijoux were transferred to a 50 ml universal and centrifuged at 230 ×g for 5 minutes to remove cell debris. The supernatant containing *C. trachomatis* EBs was added to an equal volume of 4X Sucrose Phosphate (4SP) buffer (a preservative) in a cryovial, and was stored at -80°C.

2.18 Whole Genome Sequencing (WGS)

Genomic DNA (gDNA) was isolated from *C. trachomatis* inocula (harvested from infected McCoy cells in T25 flasks; **2.17**) using the NucleoSpin® Tissue kit (*Macherey-Nagel, Germany*) (**2.4.2**). The gDNA concentration was quantified using fluorometry with Qubit® 1.0 (**2.12**). Genomic DNA (150 ng per isolate) was sent to MicrobesNG for whole genome illumina sequencing. Genomic libraries were prepared using the Nextera XT Library Prep Kit (*Illumina, USA*). MicrobesNG sequenced libraries using a 250 bp paired end protocol, and reads were adapter-trimmed using Trimmomatic (version 0.30) (Bolger, Lohse and Usadel, 2014). De novo assembly was performed using SPAdes (version 3.7) (Bankevich *et al.*, 2012). Contigs were annotated using Prokka (version 1.12) (Seemann, 2014), with the *Chlamydia* genus and a list of proteins (in fasta format) derived from the previously-annotated complete genome L2b/UCH-1/proctitis (Genbank Accession No. AM884177) as a basis for annotation.

2.19 Data Analysis

2.19.1 Visualisation and Alignment of DNA sequences

Sequence chromatograms were visualised using SnapGene® (version 5.3). This gave an idea of overall sequence quality prior to sequence alignment.

VNTR and *ompA* gene sequences were aligned using the built-in multiple sequence alignment ClustalW software on BioEdit (version 7.0.5.3). The BioEdit program was also used to identify single nucleotide polymorphisms (SNPs). Sequence base cells assigned "N" were manually checked by comparing with the chromatogram. If required, start and end bases were trimmed to remove poor-quality sequence.

2.19.2 Assignment of MLVA-*ompA* genotypes

Trimmed VNTR sequences (**2.19.1**) were analysed using BioEdit, by comparing the number of repeating mononucleotides at each VNTR locus against previously defined VNTR variants published in Pedersen *et al*. (2008), Wang *et al.* (2011), Labiran *et al.* (2017), Satoh *et al.* (2014), and Qin *et al.* (2016) (**Table 5**). A single-digit number was assigned at each VNTR locus based on the number of repeating mononucleotides. Any new VNTR sequences were assigned new VNTR types.

Alphabetical *ompA* genotypes were assigned to each *ompA* sequence by comparing the trimmed sequence data in BioEdit (**2.19.1**) against the NCBI database using the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). BLAST takes a query sequence, compares it against all sequences stored in the NCBI database, and prepares a list of results, assigning a score to each result based on the degree of sequence similarity to the query. Results with the highest scores are likely to be the most closely related to the query. The final MLVA-*ompA* genotype was designated by a 3-digit code in the order: CT1335, CT1299, and CT1291, followed by the alphabetical *ompA* genotype (e.g. 1.9.2-L2).

(A)

VNTR type code	Sequence at CT1335 locus[a]
1	GAAAAAG-**9T8A**-GCTTTTGT
2	GAAAAAGG-**10T7A**-GCTTTTGT
3	GAAAAAGG-**10T8A**-GCTTTTGT
3a	GAAAAAAG-**10T8A**-GCTTTTGT
4	GAAAAAGG-**11T7A**-GCTTTTGT
5	GAAAAAAG-**11T8A**-GCTTTTGT
6	GAAAAAGG-**12T7A**-GCTTTTGT
7	GAAAAAGG-**12T8A**-GCTTTTGT
8	GAAAAAGG-**13T7A**-GCTTTTGT
9	GAAAAAGG-**13T6A**-GCTTTTGT
11	GAAAAAAG-**7T9A**-GCTTTTGT
12	GAAAAAAG-**8T9A**-GCTTTTGT
13	GAAAAAGG-**9T9A**-GCTTTTGT
14	GAAAAAGG-**6T7A**-GCTTTTGT
15	11T9A

(B)

VNTR type code	Sequence at CT1299 locus[a]
1	TTTTTATTCT-**7C**-ATCAAA
2	TTTTTATTCT-**8C**-ATCAAA
3	TTTTTATTCT-**9C**-ATCAAA
3a	TTTTTATTCT-**9C**-T3C-ATCAAA
4	TTTTTATTCT-**10C**-ATCAAA
4a	TTTTTATTCT-**10C**-T3C-ATCAAA
4b	TTTTTATTCT-CT-**10C**-T3C-ATCAAA
5	TTTTTATTCT-**11C**-ATCAAA
5a	TTTTTATTCT-**11C**-T3C-ATCAAA
6	TTTTTATTCT-**12C**-ATCAAA
6a	TTTTTATTCT-**12C**-T3C-ATCAAA
7	TTTTTATTCT-**13C**-ATCAAA
8	TTTTTATTCT-**14C**-ATCAAA
9	TTTTTATTCT-3C2T-**6C**-ATCAAA
10	15C
11	TTTTTATTCT-**16C**-T3C-ATCAAA
12	18C
13	TTTTTATTCT-**21C**-ATCAAA

(C)

VNTR type code	Sequence at CT1291 locus[a]
1	AAAATGGTCTA-**6C**-TATTG
2	AAAATGGTCT-**8C**-TATTG
2a	AAAATGGTCTA-**8C**-TATTG
2b	AAAATAGTCTA-**8C**-TATTG
3	AAAATGGTCT-**9C**-TATTG
3b	AAAATAGTCTA-**9C**-TATTG
4	AAAATGGTCT-**10C**-TATTG
5	AAAATGGTCT-**11C**-TATTG
6	AAAATGGTCT-CT-**5C**-TATTG
7	AAAATGGTCT-**12C**-TATTG
8	AAAATGGTCT-**7C**-TATTG
8b	AAAATAGTCTA-**7C**-TATTG
9	AAAATGGTCT-**14C**-TATTG

Table 5 VNTR type codes assigned to the three VNTR loci (CT1335, CT1299, and CT1291) of the MLVA-*ompA* genotyping system (Pedersen *et al.*, 2008)**.** (A) Type codes of known CT1335 variants; (B) Type codes of known CT1299 variants; (C) Type codes of known CT1291 variants. VNTR variants were defined in Pedersen *et al.* (Pedersen *et al.*, 2008), Wang *et al.* (Wang *et al.*, 2011), Labiran *et al.* (Labiran *et al.*, 2016, 2017), Satoh *et al.* (Satoh *et al.*, 2014) and Qin *et al.* (Qin *et al.*, 2016). [a] Repeating mononucleotide sequences at each VNTR locus are shown in bold; variation in regions flanking repeats are underlined. Highlighted in blue are VNTR type codes that were identified by Labiran *et al.* (Labiran *et al.*, 2017) however the sequence flanking the repeat was not published.

2.19.3 **Phylogenetic Trees**

For phylogenetic analyses of the *ompA* gene, sequences were aligned using MEGA-X (version 10.1.8). Alignments were performed with ClustalW, following which Maximum Likelihood trees were constructed. Maximum likelihood trees can be interpreted as a hypothesis of the evolutionary history which (according to the model) would most likely have given rise to the observed data set. The maximum likelihood model selects trees and branches based on the highest probability of their formation (Szabo, O'Neill and Clarke, 2020).

The phylogeny of alignments was deduced using the Maximum Likelihood method with Tamura-Nei nucleotide substitution model with uniform rates. The Maximum Likelihood test was statistically supported by 1,000 bootstrap replications, to give the probability of each branch's formation. Bootstrapping gives the probability of the same branch being formed upon sampling each taxa multiple times. Bootstrap results are a measure of the probability that the estimated tree represents the true phylogeny (Hillis and Bull, 1993). Bootstrap proportions of ≥70% typically correspond to a ≥95% probability that the corresponding clade is real (Hillis and Bull, 1993). The out-group for phylogenetic trees was the *C. muridarum* Nigg strain (Genbank Accession no. M64171).

2.19.4 **Whole genome sequence analysis**

Sequencing reads (fastq files) of isolates were mapped against the *C. trachomatis* L2b/UCH-1/proctitis genome (Genbank accession no. AM884177), using Snippy (version 2.2) (Seemann, 2018). Read data was also mapped against the plasmid sequence for L2b/UCH-1/proctitis (Genbank Accession No. AM886279). This allowed insertions, deletions, recombination events, and single nucleotide polymorphisms (SNPs) to be identified within genome sequences. All called SNPs were checked manually using BAM files visualised in Artemis (http://sanger-pathogens.github.io/Artemis/Artemis/) (Carver *et al.*, 2012). Artemis is a free genome browser and annotation tool that allows visualisation of prokaryotic and eukaryotic genomes.

To observe the relative genetic relatedness of isolates to a reference genome, SNP data generated using Snippy (version 2.2) was combined to produce a core SNP genome using Snippy core (version 4.6.0). MEGA-X (version 10.1.8) was used to produce a phylogenetic tree based on the divergence of each isolate from the core genome, using the Generalised Time Reversible (GTR) model with 1,000 bootstraps. This phylogenetic tree was rooted on L2b/UCH-1/proctitis.

2.20 Statistical Analysis

2.20.1 Simpson's Index of Diversity

The discriminatory power of genotyping systems is defined as its ability to assign a different genotype to two unrelated strains sampled randomly from the population of a given species (Hunter and Gaston, 1988; van Belkum *et al.*, 2007). Discriminatory power is expressed as a probability and is calculated using Simpson's Index of Diversity (*D*) as described by Hunter and Gaston (Hunter and Gaston, 1988):

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} n_j(n_j - 1)$$

Whereby, *N* is the total number of samples tested, *s* is the total number of different genotypes identified, and n_j is the number of samples belonging to the *j*th type.

The discriminatory index is a value between 0 and 1. An index value of 0.95 is generally regarded as the index value for a genotyping system to have more or less "ideal" resolution (van Belkum *et al.*, 2007).

2.20.2 Typeability of genotyping systems

The typeability of a genotyping system is its ability to assign a genotype to all isolates tested by the genotyping system. It is expressed as a percentage of typeable isolates over the total number of isolates analysed (typeable and non-typeable) (van Belkum *et al.*, 2007). For example, a genotyping system that can assign a genotype to 100 isolates out of 200 isolates has a typeability of 50%.

2.21 Ethics

Ethical approval was required to obtain clinical specimens from individuals attending London GUM clinics (**Chapter 5**). This approvals process is detailed in Chapter 5. The reference numbers for the study were as follows: ERGO (45840), IRAS (254542), and REC (19/ES/0012).

CHAPTER 3 *OmpA* and *TrpA* Gene Sequence Analysis of *C. trachomatis* DNA from a Child Recently Resident in Afghanistan

Part of the work described in this chapter has been published (Mitchell *et al.*, 2021) **(Appendix 1**).

3.1 Introduction

As described in section 1.11, *ompA* genotyping is widely used to differentiate *C. trachomatis* strains on the basis of their *ompA* genotypes. *C. trachomatis* strains with *ompA* genotypes A – C cause the blinding disease trachoma (**1.6.1**), and *C. trachomatis* strains with *ompA* genotypes D - K are primarily associated with sexually transmitted urogenital tract infections (**1.6.2**); however, ocular infections can result from exposure to infected genital secretions during childbirth (neonatal conjunctivitis) (**1.7**), or via hand-to-eye spread of infected genital secretions upon sexual contact (adult inclusion conjunctivitis) (**1.6.3**). Adult inclusion conjunctivitis typically occurs in sexually active young adults between the ages of 15 - 24 years, whereas neonatal conjunctivitis typically develops between 5 – 14 days after birth (**1.7**). Rare cases of perinatally transmitted infections have been reported in children outside of the neonatal period (Oroz, 2001; Thompson, 2001). In 2001, a family cluster of *C. trachomatis* infections was reported, including a 6 year old female in whom there was no evidence of sexual abuse (Thompson, 2001). However, these cases are rare, and in non-trachoma endemic countries, chlamydial conjunctivitis in pre-pubertal children outside of the neonatal period is uncommon and is indicative of child sexual abuse (Brownell *et al.*, 2020; Mitchell *et al.*, 2021).

In this Chapter, we describe a case of conjunctivitis in a 10 year old female with a positive *C. trachomatis* diagnosis from a conjunctival swab. The child had travelled to the UK from Afghanistan a month prior to symptom onset. It was important to ascertain whether the child had been infected with a *C. trachomatis* strain associated with trachoma (*ompA* genotypes A - C), or a *C. trachomatis* strain associated with sexually-acquired chlamydial conjunctivitis (*ompA* genotypes D - K). If a urogenital *C. trachomatis ompA* genotype was detected in the conjunctival swab from the child, this would represent strong evidence of sexual abuse. As a result, the *C. trachomatis*-positive conjunctival swab collected from the 10 year old female was sent to our laboratory for *ompA* genotyping. In addition, given that polymorphisms in the *C. trachomatis* tryptophan synthase alpha (*trpA*) gene allows urogenital and trachoma-causing *C. trachomatis* strains to be distinguished (Caldwell *et al.*, 2003; Bommana *et al.*, 2021) (**1.5.1.1**), we also performed sequence analysis of the

trpA gene from conjunctival swab *C. trachomatis* DNA collected from the 10-year-old female. The hypothesis for this Chapter was that sequence variation in the *ompA* and *trpA* genes could confirm that the child had been infected with a *C. trachomatis* strain causing trachoma, as opposed to a urogenital *C. trachomatis* strain causing conjunctivitis.

3.2 Aims of this Chapter

- To determine the *ompA* genotype of *C. trachomatis* DNA extracted from a conjunctival swab collected from a 10 year old female recently resident in Afghanistan.
- To sequence and identify SNPs in the *trpA* gene of *C. trachomatis* DNA extracted from the conjunctival swab.

3.3 Methods

3.3.1 Case Presentation

A 10 year old female presented to the Ophthalmology Department at Birmingham Children's Hospital in November 2017 with a 7 month history of persistent itchy, watery left-eye and mild eyelid swelling. She had travelled by land to the UK from an urban centre in Afghanistan 8 months previously. She had been referred to the hospital by an optician when a course of prescribed chloramphenicol eye drops had been unable to resolve her symptoms. The Ophthalmologist at Birmingham Children's Hospital determined that the child had raised fleshy lesions in the inferior fornix, with mild follicular reaction of the superior tarsal conjunctiva. Sub-epithelial infiltrates were found throughout the cornea. A conjunctival swab of the left eye was taken and tested for *C. trachomatis*.

3.3.2 Testing and Subsequent Investigation

The conjunctival swab was tested for *C. trachomatis* using the Abbott RealTime *C. trachomatis* (CT)/*Neisseria gonorrhoeae* (NG) PCR assay (Abbott Molecular, US). The CT/NG assay is used for the dual detection of *C. trachomatis* plasmid DNA and *N. gonorrhoeae* genomic DNA. The conjunctival swab tested positive for *C. trachomatis* and negative for *N. gonorrhoeae*. A formal child protection investigation was commenced, undertaken by a paediatrician at Birmingham Children's Hospital with expertise in child sexual abuse. The child was interviewed alongside her parents with the aid of a translator. The child reported no inappropriate intimate contact, with no genital pain, vaginal bleeding, or discharge. Her parents reported no genital symptoms and had no prior history of *C.*

trachomatis infection. The investigation did not include a genital examination. This was decided by the medical team after considering the recent travel history, the patient symptomology, and the denial of inappropriate sexual contact. The child was treated with a 14 day course of oral erythromycin which resolved her symptoms. Following the course of antibiotics, conjunctival swabs were taken from both eyes and tested for *C. trachomatis* using the Abbott CT/NG assay. Both swabs tested negative for *C. trachomatis* and *N. gonorrhoeae*.

The initial left conjunctival swab (containing *C. trachomatis* DNA) was sent to the Department of Molecular Microbiology at the University of Southampton for further analysis. The swab arrived in lysis buffer within its original swab kit and was stored at -20°C upon arrival.

3.3.3 **Methods used to determine the *ompA* genotype**

C. trachomatis DNA was extracted from the conjunctival swab using the NucleoSpin® Tissue kit (*Macherey-Nagel, Germany*) (**2.4.2**). PCR amplification of *ompA* was carried out using primers PCTM3 and NRI following the method outlined in section 2.9.1. In total, four PCR tubes were set up, corresponding to a no-template negative control (NTC), a positive control (a transformed plasmid-less L2 *C. trachomatis* strain unlikely to be found in the general population, L2 P^-), and the template in duplicate (neat and diluted 1:8).

To confirm that the *ompA* gene had been amplified, the PCR amplicons were visualised using agarose gel electrophoresis (**Figure 12**). PCR amplicons were purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega, UK)* (**2.10**). PCR amplicons (10 ng/µL) were commercially sequenced at Source Bioscience (**2.13**).

Alignment of the swab *ompA* sequence and phylogenetic analyses were carried out using MEGA-X (version 10.1.8) software (**2.19.3**). Phylogenetic trees were constructed using the maximum likelihood method with 1,000 bootstrap replications for branch support (**2.19.3**). To allow the phylogenetic tree to be rooted, *C. muridarum* Nigg strain (Genbank Accession No. M64171) was used as an out-group in phylogenetic analyses.

3.3.4 **Methods used to determine the *trpA* gene sequence**

3.3.4.1 **PCR amplification of *trpA* and sequencing of *trpA* PCR product**

PCR amplification of *trpA* was carried out using the PCR conditions listed in Table 6. The primers trpA_F and trpA_R (*Eurogentec, Belgium*) (**Table 7**) were used to amplify the *trpA* gene. The primers were designed based on the *trpA* gene sequence of C/TW-3 using SnapGene® (version 5.3). In total,

three PCR tubes were set up, corresponding to a no-template negative control (NTC), a positive control (E/SW3; Genbank Accession No. HE601801.1), and the template (neat genomic DNA). Each tube had a total reaction volume of 20 µL, consisting of: 10 µL of Phusion Flash High-Fidelity PCR Master Mix (*Thermo Scientific™, UK*), 0.5 µM of the forward and reverse primers (*Eurogentec, Belgium*), 7 µL of UltraPure™ DNase/RNase-Free distilled H_2O *(Thermo Scientific™, UK)* and 1 µL of DNA. PCR products were loaded onto 1% agarose gels for the purpose of checking amplicon size and quality. The PCR amplicons were subsequently purified using the Wizard SV Gel and PCR Clean-Up System (*Promega, UK*) (**2.10**) prior to sequencing. *TrpA* PCR amplicons were commercially sequenced at Source Bioscience Cambridge (**2.13).**

Table 6 PCR conditions for the amplification of *trpA*

Denature		98°C	10 s
	35 cycles	98°C	2 s
Anneal		54°C	5 s
Extend		72°C	30 s
		72°C	1 min
		10°C	∞

Table 7 Primer sequences for PCR of *trpA*

Primer Name	**Primer sequence (5' - 3')**	**Amplicon size[a] (bp)**	**Nucleotide position[a]**
trpA_F	ATGAGTAAACTAACCCAAGTTTTTAAAC	758	196,300 – 196,327
trpA_R	TTATCCAGGAATAACTGTTTGTGC		197,057 – 197,034

[a] According to C/TW3, Genbank Accession No. NC_023060.1

3.3.4.2 Cloning of *trpA*

As detailed in section 3.4.2, PCR amplification of the *trpA* gene failed to yield a readable sequence, and repeat sequencing and repeat PCR of *trpA* did not improve the sequence quality. It was therefore decided to clone the *trpA* gene. The *trpA* gene was amplified by PCR using the cloning

primers trpA_XhoI_F and trpA_HindIII_R (**Table 8**). These primers added XhoI and HindIII restriction sites to the *trpA* PCR product. PCR was conducted in 50 µL reaction volumes: 25 µL of Phusion Flash High-Fidelity PCR Master Mix (*Thermo Scientific™, UK*), 2.5 µL of the forward and reverse primers (*Eurogentec, Belgium*), 18 µL of UltraPure™ DNase/RNase-Free distilled H_2O *(Thermo Scientific™, UK)* and 2 µL of DNA. PCR products were loaded onto a 1% agarose gel for the purpose of checking amplicon size and quality. *TrpA* amplicons were subsequently purified using the Wizard SV Gel and PCR Clean-Up System (*Promega, UK*) (**2.10**). The *trpA* PCR product (2 µg) was digested with HindIII and XhoI (**2.6**).

Table 8 Cloning Primer sequences for PCR of *trpA*. XhoI and HindIII restriction sites are indicated in brackets.

Primer Name	**Primer sequence (5' - 3')**
trpA_XhoI_F	GGTGGT(CTCGAG)ATGAGTAAACTAACCCAAGTTTTTAAAC
trpA_HindIII_R	(AAGCTT)GGTGGTTTATCCAGGAATAACTGTTTGTGC

Colonies of *E. coli* JM101 transformed with the pUC19 plasmid vector (**2.8**) were grown in liquid culture overnight. Plasmid DNA was isolated using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*) (**2.4.1**) (**Figure 11**). pUC19 DNA (2 µg) was digested with HindIII and SalI (**2.6**). The vector was digested with SalI and not XhoI, as pUC19 lacks a XhoI restriction site. As XhoI and SalI leave complementary overhangs, ends generated with XhoI (insert) can be directly ligated to ends generated by SalI (vector).

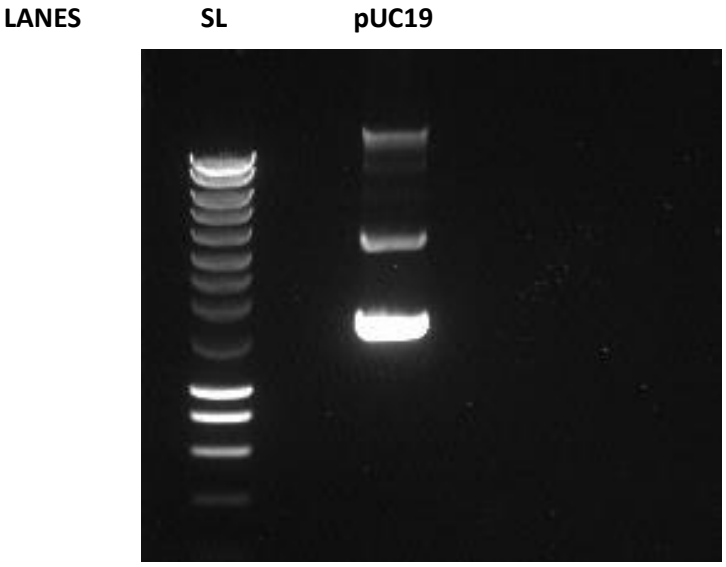


Figure 11 pUC19 DNA isolated from *E. coli* (pUC19) JM101 using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*). SL = DNA ladder Smartladder.

Digested insert DNA and pUC19 were ligated (**2.7**), and the resulting recombinant pUC19 was transformed (**2.8**) into competent *E. coli* JM101 (**2.2.2**) alongside appropriate controls. Transformed *E. coli* were plated onto X-Gal/IPTG/Ampicillin plates to allow for blue-white screening (**2.8.1**). Six white colonies and three blue colonies were selected for PCR screening, to demonstrate that *trpA* had been successfully inserted into the plasmid. These colonies were sub-cultured on a LB plate (plus Ampicillin) (**2.1.2**) overnight. Ten PCR tubes were set up corresponding to six white colonies (W1 - W6), three blue colonies (B1 - B3) and a no-template negative control (NTC). Each tube consisted of 10 µL Biomix™ Red (*Bioline, UK*), 0.5 µM forward and reverse primers (trpA_XhoI_F, and trpA_HindIII_R), 8 µL of UltraPure™ DNase/RNase-Free distilled H_2O *(Thermo Scientific™, UK)* and a scoop of the corresponding colony. The PCR conditions were as described in **Table 6**. PCR products were visualised on a 1% agarose gel (**Figure 15**). The six white colonies were purified using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*) (**Figure 16**) (**2.4.1**) and sent to Source Bioscience for sequencing using the M13F and M13R sequencing primers (**2.13**). These primers encompass the entire multiple cloning site (MCS) of pUC19.

3.4 Results

3.4.1 *OmpA* gene sequence analysis

OmpA was amplified by PCR using primers PCTM3 and NRI (**2.9**) from DNA extracted from the conjunctival swab. *OmpA* PCR products were visualised on a 1% agarose gel (**Figure 12**). A single PCR product was visible in lanes loaded with neat genomic DNA (lane 2), a 1:8 dilution of neat genomic DNA (lane 3), and the positive control (L2 P^-, lane 5). The *ompA* PCR product from lane 2 was purified and sequenced (**3.3.3**), yielding a 1,020 bp sequence. The 1,020 bp *ompA* sequence ("conjunctival swab DNA") was aligned against *ompA* of 58 *C. trachomatis* strains deposited in Genbank. The 58 *C. trachomatis* strains represented 15 *ompA* genotypes: A, B, C, D, E, F, G, H, Ia, J, K, L1, L2, L2b and L3. Figure 13 shows the maximum likelihood tree constructed from the *ompA* sequence alignment. The *C. muridarum* Nigg strain (Genbank Accession No. M64171) was used as an out-group (**Figure 13**)

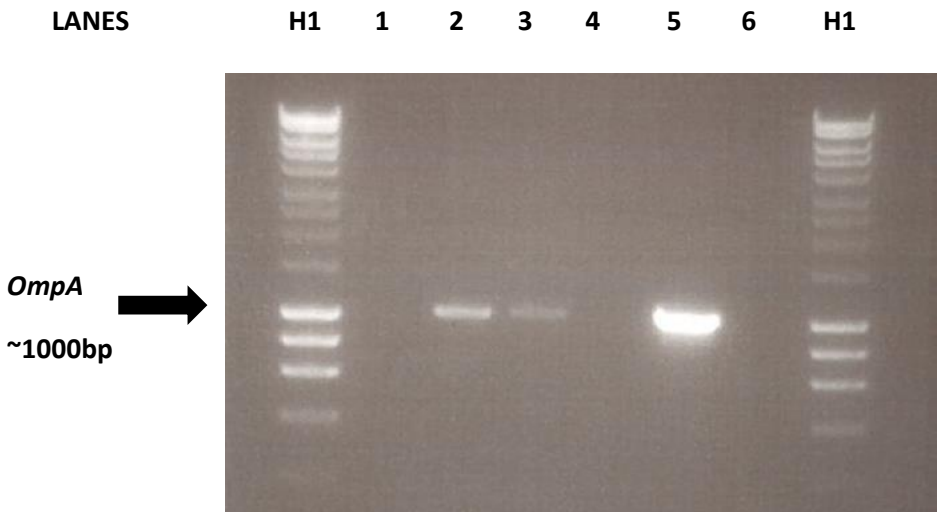


Figure 12 A 1% agarose gel showing the PCR products after amplification of *ompA* from *C. trachomatis* genomic DNA extracted from a conjunctival swab, and appropriate controls. *OmpA* was amplified by PCR using primers PCTM3 and NRI. Lanes 1 and 4 were empty wells. Lane 2 shows the products of *ompA* amplification in neat genomic DNA from the conjunctival swab. Lane 3 shows products from *ompA* amplification in a 1:8 dilution of genomic DNA from the swab. Lane 5 shows products from *ompA* amplification in the positive control (transformed plasmid-free L2 *C. trachomatis* strain, L2 P^-). Lane 6 shows products from *ompA* amplification in the no-template negative control (NTC). H1 represents the DNA ladder Hyperladder 1 (**Figure 10**).

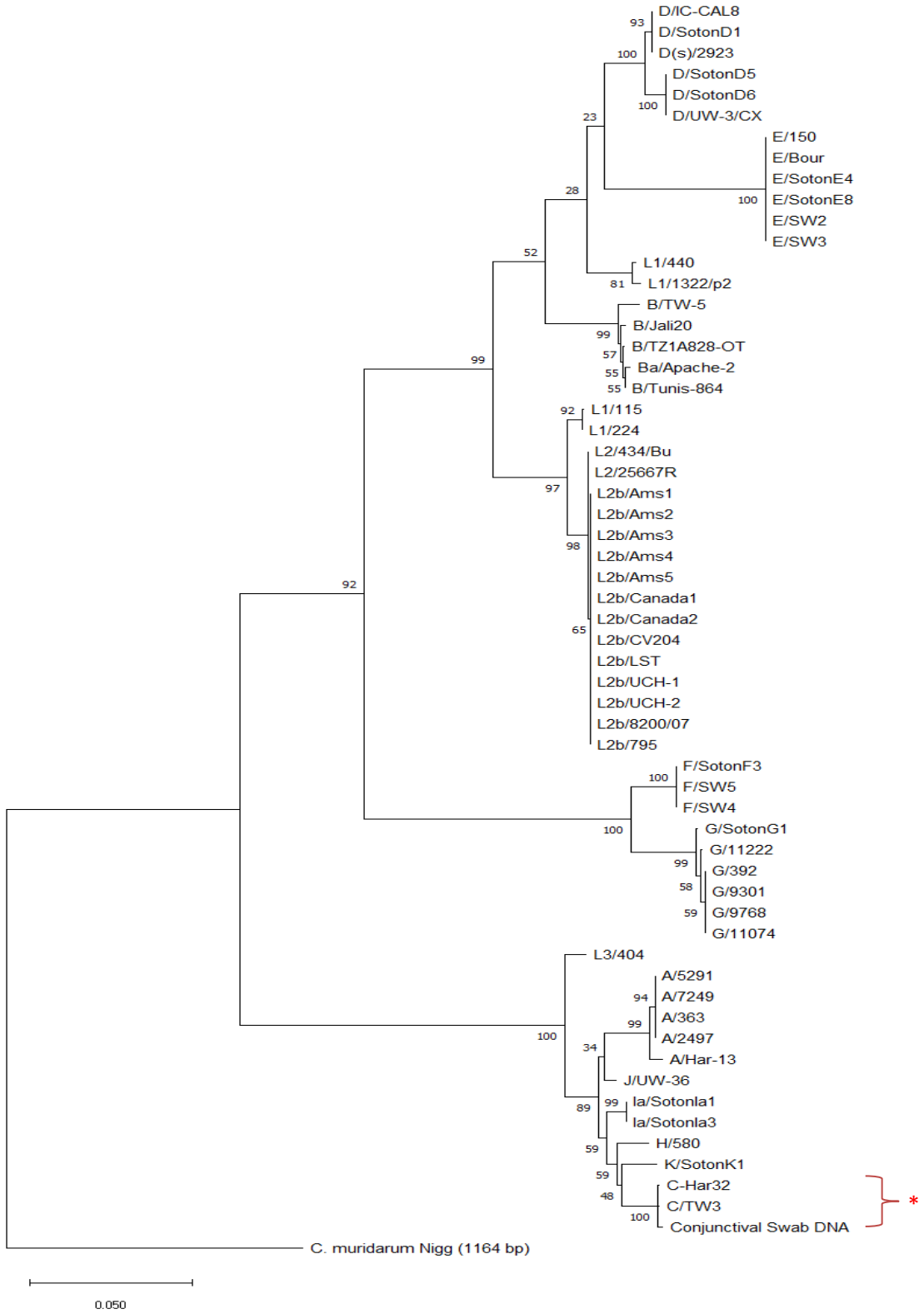


Figure 13 Maximum Likelihood Reconstruction of the Phylogeny of the *C. trachomatis ompA* nucleotide sequence from the conjunctival swab ("conjunctival swab DNA", labelled with *). The swab ompA sequence was aligned with the *ompA* of 58 *C. trachomatis* strains deposited in Genbank. The swab *ompA* sequence clustered with that of strains C/TW-3 (Genbank accession: NC_023060.1) and C-Har32 (Genbank accession: DQ064298.1) (indicated with red bracket). The tree was constructed in MEGA-X (version 10.1.8) using the maximum likelihood method with 1000 bootstrap replications for branch support. Bootstrap support for nodes on the tree is indicated. The scale bar represents the number of nucleotide substitutions per site. *C. muridarum* Nigg strain (Genbank Accession No. M64171) was used as an out-group.

Figure 13 shows that the *ompA* sequence of the swab ("conjunctival swab DNA", labelled with *) clustered with *ompA* of two *C. trachomatis* strains with *ompA* genotype C: C/TW3 (Genbank Accession No. NC_023060.1) and C/Har32 (Genbank Accession No. DQ064298.1) (indicated with a red bracket). The C/TW-3 strain was first isolated in Taiwan in 1959 from the human conjunctiva (Grayston et al., 1960), and C/Har32 was first isolated from the conjunctiva of a patient with active trachoma (McComb et al., 1967). These three *ompA* sequences comprised a well-supported clade, with a bootstrap value of 100%. The upper nodes to this clade had lower bootstrap values (59% and 48%), which indicated lower confidence in the location of the clade relative to the other sequences in the tree.

The *ompA* sequence from the swab ("conjunctival swab DNA") was compared to *ompA* of *C. trachomatis* strains C/TW3 and C/Har32. The *ompA* sequence of the C/TW3 strain (NC_023060) was used as a reference. There were three base pair changes between C/TW3 and the swab *ompA* (**Table 9**). All three changes were within variable segment 4 (VS 4) of the *ompA* gene, and all three resulted in amino acid substitutions. At amino acids 329 and 331, there were 2 single nucleotide substitutions (SNPs) within the swab *ompA* that were not seen in C/TW3, or in C/Har32 (**Table 9**). These SNPs resulted in two amino acid substitutions, K329R and S331N, respectively. The conjunctival swab *ompA* sequence also had a SNP at nucleotide position 1,033 (position relative to C/TW3), resulting in the amino acid substitution A335S. This SNP was also present in the *ompA* sequence of C/Har-32 (**Table 9**).

Table 9 *OmpA* sequence variation between the conjunctival swab and two *C. trachomatis* strains with *ompA* genotype C, C/TW3 and C/Har32. Single nucleotide polymorphisms (SNPs) and their locations are shown. The strain C/TW3 (NC_023060) was used as a reference. *OmpA* of the *C. trachomatis* strain C/Har32 (DQ064298) was also included. Letters in brackets represent the amino acid and any resulting change. VS = variable segment.

	VS 1	**VS 4**		
Nucleotide (amino acid)	**212 (71)**	**986 (329)**	**992 (331)**	**1,003 (335)**
C/TW3	A (D)	A (K)	G (S)	G (A)
C/Har32	C (D→A)	A (K)	G (S)	T (A→S)
Conjunctival Swab DNA	A (D)	G (K→R)	A (S→N)	T (A→S)

3.4.2 *TrpA* gene sequence analysis

TrpA was amplified by PCR using primers trpA_F and trpA_R (**3.3.4**) from swab genomic DNA and a positive control (E/SW3). *TrpA* PCR products were visualised on a 1% agarose gel (**Figure 14**). A single PCR product was visible in lanes loaded with neat genomic DNA (lane 2), and the positive control (lane 4). The swab *trpA* PCR product was purified (**2.10**) and sequenced (**2.13**), yielding a 720 bp sequence.

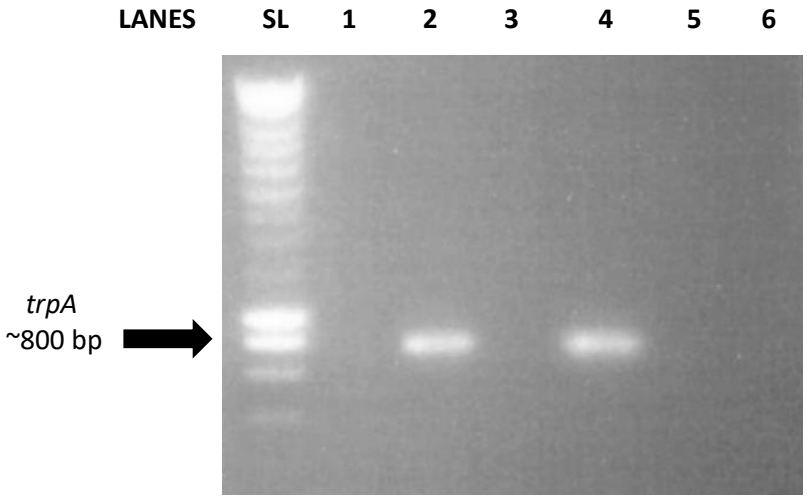


Figure 14 A 1% agarose gel showing the PCR products after amplification of *trpA* from neat genomic DNA extracted from the conjunctival swab, and appropriate controls. *TrpA* was amplified by PCR using primers trpA_F and trpA_R. Lanes 1, 3 and 5 were empty wells. Lane 2 shows the products of *trpA* amplification in neat genomic DNA from the conjunctival swab. Lane 4 shows products from *trpA* amplification in the positive control (*C. trachomatis* strain E/SW3; Genbank Accession No. HE601801). Lane 6 shows products from *trpA* amplification in the no-template negative control (NTC). SL = DNA ladder Smartladder.

The swab *trpA* forward and reverse sequence traces were visualised using SnapGene® (version 5.3). The sequence quality of both traces was poor with a high level of background, making it impossible to call peaks manually. Re-sequencing of the sample yielded the same result. A repeat PCR resulted in no improvement in sequence quality. It was therefore decided to clone the *trpA* gene using the pUC19 cloning vector following the method outlined in section 3.3.4.2. pUC19 (**2.3**) was selected as it has a multiple cloning site (MCS) located within the lacZα gene, allowing for blue-white screening of colonies transformed with recombinant plasmid (**2.8.1**). *E. coli* JM101 (**2.2**) was transformed with pUC19 containing the *trpA* insert, and spread on an X-Gal/IPTG/Ampicillin plate using the method outlined in section 3.3.4.2. Six white colonies (*E. coli* JM101 transformed with recombinant pUC19)

and three blue colonies (*E. coli* JM101 not transformed with recombinant pUC19) were selected from the plate and sub-cultured. These colonies were screened by PCR for *trpA* (**Figure 15**). An intense band (approx. 800 bp in size) was visible in lanes loaded with DNA from *E. coli* JM101 colonies transformed with pUC19 containing the *trpA* insert (W1 - W6). This indicated that the *trpA* insert was present in all six white colonies, as expected. Primer dimers were visible (<100 bp in size) in each lane loaded with sample. A faint PCR product, also approximately 800 bp in size, was visible in two out of three lanes loaded with DNA from *E. coli* JM101 colonies that appeared blue on the X-Gal/IPTG/Ampicillin plate (B1 and B2), i.e. colonies that were not transformed with pUC19 containing the *trpA* insert.

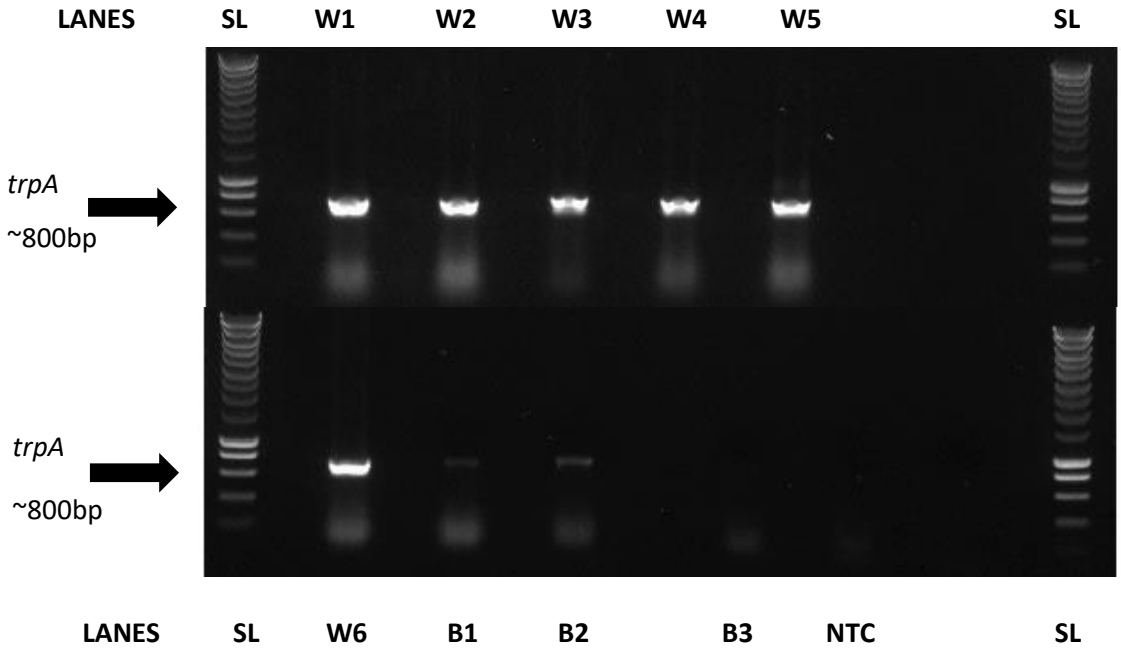


Figure 15 PCR screen of *trpA* in nine transformed *E. coli* JM101 colonies. White colonies (labelled W1 – W6) were *E. coli* JM101 colonies transformed with a recombinant pUC19 plasmid, appearing white on an X-Gal/IPTG/Ampicillin plate. Blue colonies (labelled B1 – B3) were *E. coli* JM101 colonies that were not transformed, appearing blue on an X-Gal/IPTG/Ampicillin plate. NTC refers to a no template negative control, and SL refers to the DNA ladder Smartladder.

E. coli JM101 colonies B1 and B2, and W1 - W6 from the PCR screening plate were sub-cultured onto individual X-Gal/IPTG/Ampicillin plates and incubated overnight at 37°C. Plates streaked with *E. coli* colonies B1 and B2 were blue, indicating that no insert was present. Colonies were scraped from plates streaked with W1 - W6, and pUC19 DNA was isolated using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*) (**Figure 16**). Purified recombinant pUC19 plasmids W1 - W6 were sequenced (**2.13**), and the *trpA* gene sequence was extracted and aligned against *trpA* of 50 *C. trachomatis* strains deposited in Genbank (**Appendix 2**).

As shown in Appendix 2, *trpA* from W1 – W6 contained a SNP (T10C) that was present in the reference (C/TW3) but absent from all other *C. trachomatis* strains included in the alignment. *TrpA* from W2, W3 and W4 contained the ATT deletion at nucleotide position 408 - 410 that is characteristic of trachoma-causing *C. trachomatis* strains (*ompA* genotypes A – C). However, *trpA* from W1, W5 and W6 contained ATT at position 408 - 410. This is characteristic of urogenital *C. trachomatis* strains (*ompA* genotypes D – K). At nucleotide position 531, *trpA* from W2 and W3 had a single nucleotide deletion (TAC-AA) resulting in a frame shift. W1, W5 and W6 had a single nucleotide substitution (TATCAA), resulting in the amino acid substitution YQ. W4 had a single nucleotide substitution (TACCAA), also resulting in the amino acid substitution YQ (**Appendix 2**).

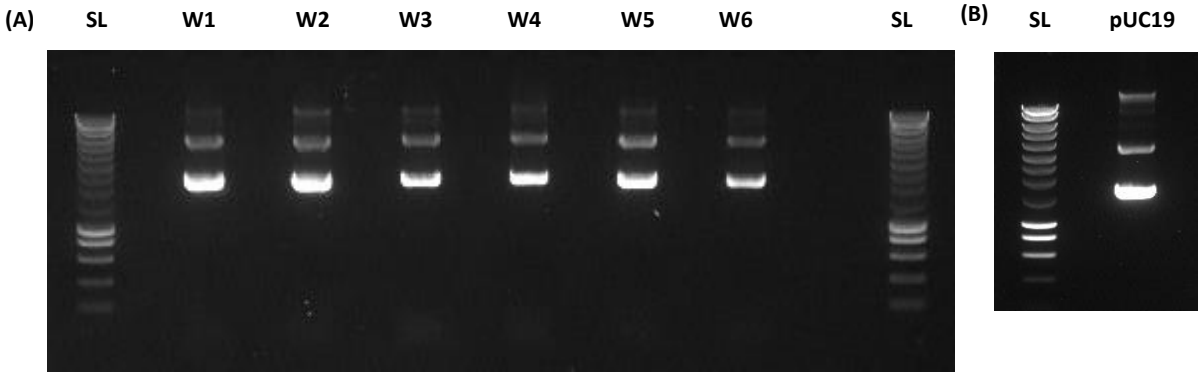


Figure 16 Two 1% agarose gels showing (A) Purified recombinant pUC19 vectors containing the *trpA* gene insert (isolated from six transformed *E. coli* JM101 colonies; labelled W1 – W6); (B) Purified non-recombinant pUC19 vector (i.e. not containing *trpA* insert) for comparison. Purified plasmid DNA was isolated using the Wizard® *Plus* SV Minipreps DNA Purification System (*Promega, UK*). SL refers to the DNA ladder Smartladder.

3.5 Discussion

As discussed in 3.1, *C. trachomatis* infections can manifest in the eye in the form of trachoma, neonatal conjunctivitis, and adult inclusion conjunctivitis (Giffard *et al.*, 2016).

Trachoma is a blinding disease that is caused by *C. trachomatis* strains with *ompA* genotypes A – C. Trachoma is endemic across rural regions of Africa, Asia, the Middle East and Australia (World Health Organisation, 2020b). It is estimated that 137 million people live in trachoma endemic regions and are at risk of blindness from trachoma (World Health Organisation, 2020b). Active disease is most frequently detected amongst children aged 3 – 6 years and is characterised by conjunctivitis. Repeat infections can result in the build-up of scar tissue in the eyelid that can worsen and cause the eyelids to turn inwards and the eyelashes to scrape against the cornea, eventually leading to blindness. *C. trachomatis* strains with *ompA* genotypes D - K can also cause ocular infections. These include neonatal chlamydial conjunctivitis, caused by the exposure of a neonate to infected genital secretions as it passes along the birth canal; and adult inclusion conjunctivitis, caused by the hand-to-eye spread of infected genital secretions (Kestelyn, 2008). Whilst neonatal conjunctivitis typically manifests in neonates between 5 and 14 days after birth (Zar, 2005a); adult inclusion conjunctivitis is frequently reported amongst sexually active young adults aged 15 to 24 years (Kestelyn, 2008).

In non-trachoma endemic countries, chlamydial conjunctivitis is rare in children outside of the neonatal period; and a positive *C. trachomatis* diagnosis in a child is highly suggestive of child sexual abuse (Ingram *et al.*, 1992). In this Chapter, we detailed the laboratory investigation of a *C. trachomatis*-positive conjunctival swab collected from a 10 year old female with symptoms of conjunctivitis. The child had recently immigrated to the UK from Afghanistan. The *ompA* gene was sequenced to determine the *ompA* genotype of the DNA extracted from the swab. The rationale for this was to determine whether the child had been infected with a trachoma-causing *C. trachomatis* strain (*ompA* genotypes A - C), or a *C. trachomatis* strain associated with sexually transmitted adult inclusion conjunctivitis (*ompA* genotypes D - K).

Phylogenetic analysis of *ompA* showed that the swab *ompA* sequence clustered with *ompA* of two *ompA* genotype C strains - C/TW3 and C/Har32 (**Figure 13**). These three *ompA* sequences formed a well-supported clade with a bootstrap value of 100%. Strains with *ompA* genotype C are associated with trachoma, and there have been no reports of strains with this genotype causing urogenital disease (Andersson *et al.*, 2016; Hadfield *et al.*, 2017). The child had recently immigrated to the UK from Afghanistan. Whilst the trachoma endemicity status of Afghanistan is unknown due to regional instability preventing epidemiological data from being collected; trachoma has reached endemic levels in countries bordering Afghanistan, including Pakistan. Therefore it is likely that trachoma is

also endemic in Afghanistan, and this would explain the infection of the child with a *C. trachomatis* strain associated with trachoma (*ompA* genotype C). The decision to open a sexual abuse investigation was justified, given the low incidence of chlamydial conjunctivitis outside of the neonatal period in non-trachoma endemic regions. However, a key limitation of the child abuse investigation was that no genital examination was conducted on the child, and no endocervical swab specimens (or urine sample) were collected.

Sequence analysis of the swab *ompA* gene (**Table 9**) revealed two single nucleotide polymorphisms (SNPs) within variable segment 4 (VS 4) that resulted in the amino acid substitutions K329R and S331N, that were not present in the other *C. trachomatis ompA* genotype C strains. A third SNP was also present in the conjunctival swab DNA sample resulting in the amino acid substitution A335S, however this SNP was shared by C/Har-32. Whilst the first two SNP mutations in the swab *ompA* gene were not observed in the other two *C. trachomatis ompA* genotype C strains, care must be taken when interpreting these findings as truly novel. A limitation of this analysis is that only two *ompA* genotype C strains were included, which may not fully encapsulate the diversity that exists within the *ompA* gene in these strains. It is unsurprising that the three SNPs identified within the swab *ompA* were located within variable segment 4. It has been suggested that variation in MOMP may have arisen due to immune pressure. A study by Brunelle and Sensabaugh (2012) showed that there is an increase in nucleotide substitutions within *ompA* and its surrounding loci. There was also an increase noted in the rate of amino acid replacements in the variable domains of *ompA* compared to the constant domains. This is largely expected given that the variable domains are surface-exposed; hence amino acid substitutions in these regions result in variation in the MOMP epitopes that act as the antigenic targets of the host immune response. It was suggested that these substitutions produce a high degree of variability that can promote host immune evasion.

Previous studies (Fehlner-Gardiner *et al.*, 2002; Caldwell *et al.*, 2003; Bommana *et al.*, 2021) have demonstrated that trachoma-causing and urogenital *C. trachomatis* strains can be distinguished on the basis of polymorphisms in the tryptophan synthase alpha subunit (*trpA*) gene. Trachoma-causing strains (and not urogenital strains) have an ATT deletion at nucleotide positions 408 - 410, in addition to a deletion at nucleotide position 531 that results in a frame shift and the expression of a truncated TrpA polypeptide. Urogenital strains have two SNPs at nucleotide position 531 that result in amino acid substitutions that vary amongst urogenital strains: YE in LGV strains; CQ in D, E and K strains; and YQ in F - J strains. Whilst both urogenital and ocular strains are unable to produce indole in the alpha reaction; genital but not ocular strains of *C. trachomatis* are rescued under conditions of tryptophan starvation by the addition of exogenous indole (Fehlner-Gardiner *et al.*, 2002). It has been suggested that genital strains have retained the full-length tryptophan synthase so that they may still be able to synthesise tryptophan using exogenous indole that is likely sourced

from other bacterial colonisers of the genital tract (**1.5.1.1**). The original explanation as to why ocular strains have not retained an intact tryptophan synthase operon, was that the ocular environment was sterile, lacking colonisers to provide a source of indole, hence there was no need to maintain a functional tryptophan synthase (Caldwell *et al.*, 2003). More recent evidence has shown this explanation to be incorrect. The ocular environment is not indeed sterile (Willcox et al., 2013; Garza et al., 2021). However, the bacterial species isolated from the ocular environment are non-indole producing (Willcox et al., 2013). A more likely explanation as to why ocular strains have a non-functional tryptophan synthase operon, is that in the absence of indole, tryptophan synthase begins to deaminate serine to pyruvate and ammonia, the latter of which is bactericidal. This was only shown to occur when an intact tryptophan synthase was expressed (Sherchand and Aiyar, 2019). In this way, the TrpA mutations in ocular strains can be seen as protective, providing a clear survival advantage to the pathogen. Ammonia generation may indeed act as a negative selection pressure against maintaining a functional tryptophan synthase enzyme in the indole-free environment in the eye (Sherchand and Aiyar, 2019; Banerjee and Nelson, 2019).

Sequence analysis of the *swab* trpA gene was conducted to confirm that the child had been infected with a trachoma-causing strain. However, PCR amplification of the *trpA* gene failed to yield a good-quality sequence, with a high level of background visible in the sequence traces. This meant that peaks could not be manually assigned and a consensus sequence could not be determined. Re-sequencing of the PCR product yielded the same result, thus a sequencing error was ruled out. A repeat *trpA* PCR failed to improve sequence quality. A new strategy to sequence the *trpA* gene was adopted – the cloning of *trpA* using the pUC19 cloning vector. This vector was selected as it has a lacZα gene that is located within the multiple cloning site, enabling blue-white screening of transformed *E. coli* colonies. pUC19 DNA (containing the *trpA* insert) from six transformed *E. coli* JM101 colonies (W1 - W6) was purified and sequenced. The sequencing results were inconsistent – *trpA* from three plasmids (W2, W3 and W4) contained the ATT deletion at nucleotide position 408 - 410 that is characteristic of trachoma-causing *C. trachomatis* strains. However, *trpA* from W1, W5 and W6 contained ATT at the same position – this is characteristic of urogenital *C. trachomatis* strains. Further, *trpA* from W2 and W3 contained the single nucleotide deletion at nucleotide 531 that is characteristic of trachoma-causing strains; however, W1, W5 and W6 contained two SNPs at position 531 resulting in the amino acid substitution YQ; and W4 contained two SNPs also resulting in the amino acid substitution YQ. The amino acid substitution resulting in YQ is characteristic of urogenital *C. trachomatis* strains with *ompA* genotypes F - J. These results are conflicting – all six *trpA* sequences should have been identical, given that the insert (the *trpA* gene) was the same. There are a few possible explanations. Firstly, the child could have been infected with two *C. trachomatis* strains – a urogenital strain and a trachoma-causing strain – that were both isolated

when the colonies were plated. This explanation is unlikely, given that the *ompA* sequencing traces were excellent quality and were not mixed traces, which would have been indicative of a mixed infection. A second explanation could be that the child was infected with a recombinant *C. trachomatis* strain with a trachoma-causing *ompA* genotype and a urogenital *trpA* gene. Trachoma strains that are entirely "urogenital" aside from the *ompA* gene have been previously reported in Australian Aboriginal populations (Andersson *et al.*, 2016). In this case, however, this explanation is not possible. Recombinant strains would likely contain a hybrid *trpA* sequence, and not separate urogenital *trpA* and trachoma-causing *trpA* sequences. The most likely explanation is contamination of the insert *trpA* with *C. trachomatis* DNA from a urogenital strain. This would explain the background in the *trpA* sequence traces, and the fact that re-sequencing and re-PCR of the swab *trpA* did not improve sequence quality, and the isolation of pUC19 containing urogenital *trpA* sequences. However, all six *trpA* sequences contained the same SNP (a C) at nucleotide position 10 (**Appendix 2**). This SNP was only otherwise present in *trpA* of C/TW3 in the sequence alignment. Given that no urogenital strains contained this SNP, this reduces the likelihood of contamination.

3.6 Conclusions

This Chapter identified that the child had been infected with a *C. trachomatis* strain that was most closely related to strains with *ompA* genotype C. Strains with *ompA* genotype C are associated with trachoma, and there have been no reports of strains with this genotype causing urogenital disease (Andersson *et al.*, 2016; Hadfield *et al.*, 2017). The swab *ompA* contained two novel SNPs resulting in two amino acid substitutions within variable segment 4 of the *ompA* gene.

Sequence analysis of the tryptophan synthase alpha (*trpA*) gene proved inconclusive, and as a result, the trachoma-causing nature of the *C. trachomatis* strain could not be confirmed by sequencing of the *trpA* gene. Ultimately, the child was treated with oral erythromycin which resolved her symptoms, and subsequent conjunctival swabs tested negative for *C. trachomatis*.

Final diagnoses must take into account the full investigation, including patient history, symptomology, as well as molecular diagnostic evidence. On the basis of the *ompA* genotyping results, and the fact that the child immigrated to the UK from Afghanistan (a suspected trachoma endemic country); we conclude that it is highly likely that the child had trachoma caused by a hitherto un-identified trachoma strain.

CHAPTER 4 An Evaluation of High-Resolution Genotyping Systems for *C. trachomatis*

4.1 Introduction

For decades, sequence analysis of the *ompA* gene has been used to differentiate between *C. trachomatis* strains (Dean *et al.*, 1992) (**1.11**). However, whilst *ompA* genotyping is informative and is used globally for *C. trachomatis* strain classification; there are limitations to this technique. The *ompA* gene has been shown to be a mutational hotspot within the *C. trachomatis* genome (Brunelle and Sensabaugh, 2012), and recombination within this gene has been extensively reported (Gomes *et al.*, 2007; Jeffrey *et al.*, 2010; Joseph *et al.*, 2011; Somboonna *et al.*, 2011; Brunelle and Sensabaugh, 2012; Harris *et al.*, 2012; Borges *et al.*, 2019). In addition, typing systems that utilise a single marker provide little insight into the diversity across the rest of the genome. As a result, in recent years genotyping systems targeting multiple loci have been developed for *C. trachomatis*: multi-locus sequence typing (MLST) schemes (Pannekoek *et al.*, 2008; Dean *et al.*, 2009), multi-sequence typing (MST) schemes (Klint *et al.*, 2007), and multi-locus variable number tandem repeat (VNTR) analysis (MLVA) schemes (Pedersen *et al.*, 2008; Peuchant *et al.*, 2012) (**1.11**). The applications, advantages and disadvantages, and uses of each genotyping system are discussed in section 1.11.

Despite the existence of several multiple-locus genotyping systems for *C. trachomatis*; there is not one single standardised genotyping system. It is left to the researcher to determine the genotyping system that would be most suitable for application in a given genotyping study.

The aim of this Chapter was to evaluate several multi-locus typing systems for *C. trachomatis*, including the two MLST schemes (Pannekoek *et al.*, 2008; Dean *et al.*, 2009); the MST scheme (Klint *et al.*, 2007), and the MLVA-*ompA* scheme *(Pedersen et al.*, 2008), in order to assess which genotyping system would be most suitable to use to conduct a genotyping survey of LGV *C. trachomatis* strains from London (**Chapter 5**). To facilitate this, we obtained whole-genome data from LGV strains deposited in NCBI Genome, and conducted an *in silico* genotyping survey on these strains, determining the sequence types (STs) and MLVA-*ompA* genotypes for each strain. The typeability and discriminatory capacity of each scheme was assessed (**2.20**). These are parameters that are widely used to assess the ability of genotyping systems to assign genotypes, and the ability of genotyping systems to discriminate between un-related strains, respectively. A field-test of the genotyping scheme determined to be most well-suited for application in Chapter 5 was

subsequently undertaken to assess the practicability of the system for genotyping clinical specimens (i.e. ease of use, ease of assignation of genotypes etc.).

4.2 Aims and Objectives

- To evaluate several multi-locus genotyping schemes for *C. trachomatis* in order to determine which scheme would be most suitable for a genotyping survey of LGV strains collected from London-based GUM clinics (**Chapter 5**):
    - To apply the MST (Klint *et al.*, 2007), MLST (Pannekoek *et al.*, 2008), MLST (Dean *et al.*, 2009), and MLVA-*ompA* (Pedersen *et al.,* 2008) genotyping schemes to public whole genome data from LGV strains of *C. trachomatis.*
    - To determine the MST and MLST sequence types (STs) and MLVA-*ompA* genotypes for each LGV strain.
    - To calculate the discriminatory capacity for each genotyping scheme using the Hunter-Gaston modification of Simpson's Index of Diversity (van Belkum *et al.*, 2007).
    - To calculate the typeability of each genotyping scheme by determining the percentage of strains assigned STs or MLVA-*ompA* genotypes, out of the total number of strains included in the analysis.
- To consider the practicability and reproducibility of the most appropriate scheme for genotyping LGV strains of *C. trachomatis*:
    - To undertake a field-test of this genotyping scheme (Pedersen *et al.*, 2008) using DNA extracted from an LGV rectal isolate obtained from an MSM from Brighton.

4.3 Methods

4.3.1 Data Retrieval

Whole genome data was obtained from NCBI Genome (https://www.ncbi.nlm.nih.gov/genome/). The keywords "*Chlamydia trachomatis*" were inputted to produce a list of *Chlamydia trachomatis* genomes. Descriptive data was available for strains, including geographical origin, source, and collection date. As this Chapter aims to determine the genotyping system that would be best suited for genotyping LGV strains collected from London (**Chapter 5**), it was decided to evaluate the four genotyping systems using exclusively LGV strains deposited on NCBI Genome. For LGV strains with completely assembled genomes, the chromosome accession number was selected and the .fasta file was downloaded. For LGV strains with in-completely assembled genomes that had been

sequenced by whole genome shotgun strategy, the Whole Genome Shotgun Accession Number was selected and the .fasta file was downloaded.

Four genotyping schemes were evaluated in this Chapter: the Dean *et al.* MLST scheme (2009), the Pannekoek *et al.* MLST scheme (2008), the Klint *et al.* MST scheme (2007), and the Pedersen *et al.* MLVA-*ompA* genotyping scheme (2008). The decision was made to exclude the Peuchant *et al.* (2012) MLVA scheme from this Chapter, as it has not been independently validated and has not been applied outside of the original publication. The four genotyping schemes selected for evaluation in this Chapter have been widely used in epidemiological studies worldwide.

The first MLST scheme (Pannekoek *et al.*, 2008) targets the housekeeping genes *gatA*, *oppA_3*, *hflX*, *gidA*, *enoA*, *hemN,* and *fumC*. The second MLST scheme (Dean et al., 2009) targets the housekeeping genes *glyA*, *mdhC*, *pdhA, yhbG, pykF, lysS,* and *leuS*. The MST scheme (Klint *et al.*, 2007) targets the five hypervariable regions *CT058, CT144, CT172, hctB* and *pbpB*). The MLVA-*ompA* genotyping system (Pedersen et al., 2008) targets three VNTR loci, CT1335, CT1299, CT1291, in addition to the *ompA* gene. The molecular markers included in each of the four genotyping systems are listed in Table 10. The nucleotide positions and functions of each molecular marker are also included in Table 10.

The primers targeting each molecular marker are provided in Table 11. For each LGV strain, the .fasta file was opened using Artemis (version 16.0.0) (Carver *et al.*, 2012). Using the Artemis "Navigator" tool, the primer sequences were inputted to locate each molecular marker.

The website (https://pubmlst.org/) hosts an open-access database for *C. trachomatis*. This database contains *C. trachomatis* sequence data for all markers comprising the MLST schemes (Pannekoek *et al.*, 2008; Dean et al., 2009), and the MST scheme (Klint *et al.*, 2007). A sequence can be inputted and compared against existing marker sequences in the database, giving a profile for each locus. Sequence types (STs) were subsequently determined by inputting the allelic profile into https://pubmlst.org/biosdb?db=pubmlst_chlamydiales_seqdef. To assign MLVA-*ompA* genotypes to the LGV strains in this Chapter, the VNTR and *ompA* primer sequences were inputted into the Artemis "Navigator" tool to locate each sequence. A type was assigned at each VNTR locus according to the method outlined in **2.19.2**.

4.3.2 Statistical Analysis

As discussed in section 4.1, the parameters that were assessed in this Chapter to evaluate each genotyping scheme were: (1) the discriminatory power (as measured by Simpson's Index of Diversity) (**2.20.1**), and (2) the typeability of each scheme (**2.20.2**).

4.3.2.1 Discriminatory Power

The discriminatory power of a genotyping system is its ability to assign a different genotype to two unrelated strains sampled randomly from the population of a given species. It is expressed as a probability using Hunter and Gaston's modification of Simpson's index of diversity (Hunter and Gaston, 1988). The formula used to define Simpson's index of diversity, *D*, is:

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} n_j(n_j - 1)$$

where *N* is the total number of samples, *S* is the total number of genotypes identified, and n_j is the number of samples belonging to the *j*th type.

4.3.2.2 Typeability

The typeability of a genotyping scheme is its ability to assign a genotype to all samples. It is calculated by dividing the number of strains assigned STs or MLVA-*ompA* genotypes, by the total number of strains included in the analysis. Typeability is expressed as a percentage.

Table 10 Description of the molecular markers that comprise the four genotyping schemes evaluated in this Chapter.

Scheme	Molecular Marker	Name	Nucleotide position* (5'->3')	Length (bp)*	Function
MLST (Pannekoek *et al.*, 2008)	*gatA*	Glutamyl-tRNA (Gln) amidotransferase subunit A	2,108 - 3,583	1,476	Allows the formation of correctly charged Gln-tRNA through the transamidation of mis-acylated Glu-tRNA in organisms lacking glutaminyl-tRNA synthetase
	oppA_3	Oligopeptide Binding Protein	222,438 - 223,994	1,557	Binding and transport of peptides and peptidoglycans
	hflX	GTP-binding Protein	433,641 - 432,298	1,344	GTPase that associates with the 50S ribosomal subunit and may have a role during protein synthesis
	gidA	Glucose-Inhibited Division protein A	576,941 - 578,773	1,833	NAD-binding protein involved in the addition of a carboxymethylaminomethyl (cmnm) group at the wobble position (U34) of certain tRNAs
	enoA	Enolase	661,850 - 663,124	1,275	Catalyses the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP) – the final step of glycolysis.
	hemN	Coproporphyrinogen-III oxidase	60,110 - 58,974	1,137	Involved in step 1 of the pathway that synthesises protoporyphrinogen-IX from coproporphyrinogen-III.
	fumC	Fumarate hydratase	1,004,550 - 1,005,941	1,392	Catalyses the conversion of fumarate to L-malate.
MLST (Dean *et al.*, 2009)	*glyA*	Serine hydroxymethyltransferase	502,443 - 500,950	1,494	Catalyses the interconversion of serine and glycine with tetrahydrofolate (THF) serving as the one-carbon carrier.
	mdhC	Malate dehydrogenase	430,162 - 429,182	981	Catalyses the oxidation of malate to oxaloacetate.
	pdhA	Pyruvate dehydrogenase E1 component subunit α	273,563 - 274,585	1,023	Catalyses the conversion of pyruvate to acetyl-CoA and CO_2
	yhbG	ABC transporter ATP-binding protein	751,062 - 750,343	720	Binding and transport; ATP-binding cassette (ABC) transport system
	pykF	Pyruvate Kinase	374,217 - 375,674	1,458	Involved in step 5 of the pathway synthesising pyruvate from D-glyceraldehyde 3-phosphate

	lysS	Lysine-tRNA ligase	916,214 - 917,794	1,581	Translation; Lysine biosynthesis; aminoacyl-tRNA biosynthesis
	leuS	Leucine-tRNA ligase	238,225 - 235,766	2,460	Translation; aminoacyl-tRNA biosynthesis
MST (Klint *et al.*, 2007)	*CT058*		68,528 - 67,425	1,104	Encodes a hypothetical protein
	CT144		160,694 - 161,551	858	Encodes a hypothetical protein
	CT172		195,582 - 195,091	492	Encodes a hypothetical protein
	hctB	Histone-like protein	51,504 - 52,115	612	Regulation of chromatin structure and gene expression
	pbpB	Penicillin-binding protein	780,668 - 783,910	3,243	Encodes a penicillin-binding protein
MLVA-*ompA* (Pedersen *et al.*, 2008)	CT1335		737,225 - 737,377	153	VNTR[a] located within DNA topoisomerase I/SWI domain fusion protein
	CT1299		291,758 - 291,945	188	VNTR[a] located in a non-coding region
	CT1291		195,536 - 195-760	225	VNTR[a] located within hypothetical protein CT172.1
	ompA	Outer membrane protein A	780,060 - 778,879	1,182	Encodes the major outer membrane protein (MOMP)

*According to D/UW-3/Cx, Genbank accession number NC_000117; [a] VNTR refers to Variable Number Tandem Repeats.

Table 11 Primers for MLST (Pannekoek *et al.*, 2008), MLST (Dean *et al.*, 2009), MST (Klint *et al.*, 2007), and MLVA-*ompA* (Pedersen *et al*., 2008) schemes.

Scheme	Molecular Marker	Primer Name	Primer Sequence (5'→3')	Amplicon size*	Nucleotide position*
MLST (Pannekoek *et al.*, 2008)	*gatA*	gatA_F	GCTTTAGAATTARSARAWGCT	512	2,123 - 2,143
		gatA_R	GATCCTCCGGTATCYGATCC		2,634 - 2,615
	oppA_3	oppA_3_F	ATGCGCAAGATATCAGTGGG	587	222,438 - 222,457
		oppA_3_R	AAAGCTCCRSTWGMTATMGGWAG		223,024 - 223,002
	hflX	hflX_F	GCTTCTARAGTACTTTTAAATG	512	432,737 - 432,758
		hflX_R	ATTTTRGAAATYTTTKCSAGYCG		433,248 - 433,226
	gidA	gidA_F	GGAGTCWCTACWAAAGAAGG	560	577,352 - 577,371
		gidA_R	TCGTAYTGYACATCRAAAGG		577,911 - 577,892
	enoA	enoA_F	CCTATGATGAATCTKATCAATGG	431	662,288 - 662,310
		enoA_R	TCTTCTTCRGCWAGMCCATCT		662,718 - 662,698
	hemN	hemN_F	AGATCTTCTTCWGGRGGWAGAGA	491	867,799 - 867,821
		hemN_R	TTCYTTCAKAACSTAGGTTTT		868,289 - 868,269
	fumC	fumC_F	ATTAAAAAATGTGCTGCT	572	1,004,691-1,004,708
		fumC_R	CCTTCAGGAACATTYAACCC		1,005,262-1,005,243
	Where R = A or G; S = G or C; W = A or T; Y = C or T; M = A or C; K = G or T				
MLST (Dean *et al.*, 2009)	*glyA*	FglyA	GAAGACTGTGGCGCTGTTTTATGG	679	501,702 - 501,679
		RglyA	CTTCCTGAGCGATCCCTTCTGAC		501,024 - 501,046
	mdhC	FmdhC	GGAGATGTTTTTGGCCTTGATTGT	616	430,081 - 430,058
		RmdhC	CGATTACTGCACTACCACGACTCT		429,466 - 429,489
	pdhA	FpdhA	CTACAGAAGCCCGAGTTTTT	708	273,600 - 273,619
		RpdhA	CTGTTTGTTGCATGTGGTGATAAG		274,307 - 274,284
	yhbG	FyhbG	TCAAGTCAATGCAGGAGAAAT	629	750,991 - 750,971
		RyhbG	GATAGTGTTGACGTACCATAGGAT		750,363 - 750,386
	pykF	FpykF	ATCTTATCGCTGCTTCGTT	651	374,773 - 374,791
		RpykF	CAGCAATAATAGGGAGATA		375,423 - 375,405
	lysS	FlysS	GAAGGAATCGATAGAACGCATAAT	746	917,030 - 917,053
		RlysS	ATACGCCGCATAACAGGGAAAAAC		917,775 - 917,752
	leuS	FleuS	TCCCTTGGTCGATCTCCTCAC	713	237,443 - 237,423
		RleuS	GGGCATCGCAAAAACGTAAATAGT		236,731 - 236,754
MST (Klint *et al.*, 2007)	*CT058*	CT222F	CTTTTCTGAGGCTGAGTATGATTT	1,499	68,713 - 68,690
		CT1678R	CCGATTCTTACTGGGAGGGT		67,215 - 67,234
	CT144	CT144:248F	ATGATTAACGTGATTTGGTTTCCTT	799	160,641 - 160,665

		CT144:1046R	GCGCACCAAAACATAGGTACT		161,439 - 161,419
	CT172	Four268F	CCGTAGTAATGGGTGAGGGA	380	195,680 - 195,661
		Four610R	CGTCATTGCTTGCTCGGCTT		195,301 - 195,320
	hctB	Hctb39F	CTCGAAGACAATCCAGTAGCAT	796	51,219 - 51,240
		Hctb794R	CACCAGAAGCAGCTACACGT		52,014 - 51,995
	pbpB	pbpB1F	TATATGAAAAGAAAACGACGCACC	2,366	780,665 - 780,688
		pbpB2366R	TGGTCAGAAAGATGCTGCACA		783,030 - 783,010
MLVA-*ompA* (Pedersen *et al.*, 2008)	CT1335	CT1335_F	TCATAAAAGTTAAATGAAGAGGGACT	153	737,225 – 737,250
		CT1335_R	TAATCTTGGCTGGGGATTCA		737,377 – 737,358
	CT1299	CT1299_F	TTGTGTAAAGAGGGTCTATCTCCA	188	291,758 – 291,781
		CT1299_R	AAGTCCACGTTGTCATTGTACG		291,945 – 291,924
	CT1291	CT1291_F	GCCAAGAAAAACATGCTGGT	225	195,536 – 195,555
		CT1291_R	AGGATATTTCCCTCAGTTATTCG		195,760 – 195,738
	ompA	PCTM3	TCCTTGCAAGCTCTGCCTGTGGGGAATCCT	1,019	779,977 – 780,006
		NR1	CCGCAAGATTTTCTAGATTTC		779,008 – 778,988

*According to D/UW-3/Cx, Genbank accession number NC_000117.

Upon application of the aforementioned methods to the LGV strain genomes deposited in NCBI Genome, it was concluded (**4.4**) that the MLVA-*ompA* genotyping system (Pedersen *et al.*, 2008) was best suited for the genotyping survey of LGV strains from London (**Chapter 5**). Therefore, the MLVA-*ompA* genotyping system was deployed in the following field test in order to assure its suitability for genotyping clinical specimens:

4.3.3 Field-Test of the MLVA-*ompA* genotyping system

4.3.3.1 Bri038 Rectal Isolate

The rectal isolate that was genotyped in this Chapter as a field-test of the MLVA-*ompA* genotyping system was Bri038. Bri038 is an LGV isolate that was obtained using a rectal swab from an MSM from Brighton in a study by Labiran *et al.* (2016). This isolate, along with two other LGV isolates, were whole genome sequenced in Chapter 6 of this thesis. The details of Bri038 isolate collection, growth in cell culture, and genomic DNA extraction are provided in Chapter 6.

The rationale for the field-test was to assess the practicability of the MLVA-*ompA* genotyping system (i.e. its ease of use, the ease of assignation of genotypes) and also to assess the reproducibility of the method (i.e. the ability to assign the same type upon repeated testing). We selected to MLVA-*ompA* genotype the Bri038 isolate in this Chapter, as the MLVA-*ompA* genotype for this isolate had previously been determined by Labiran *et al.* (2016). As a result, we could compare the genotype assigned in this Chapter to that obtained by Labiran.

The study by Labiran *et al.* (2016) previously determined the MLVA-*ompA* genotype of Bri038 as 1.9.2 – L2b. This was achieved by using the MLVA-*ompA* primers listed in Table 12 and the PCR conditions described in section 2.9.1. For the field-test of the MLVA-*ompA* genotyping system in this Chapter, it was decided to repeat this process and confirm the MLVA-*ompA* genotype of Bri038 using genomic DNA extracted from Bri038 in Chapter 6.

4.3.2.1 MLVA-*ompA* genotyping of Bri038

Genomic DNA extracted from Bri038 (**Chapter 6**) was PCR amplified using the MLVA-*ompA* primers previously used by Labiran *et al.* (2014) (**Table 12**). A positive control, SotonK1 (*ompA* genotype K; Genbank Accession no. HE601794.1), and a no-template negative control (NTC) were employed. PCR amplicons were visualised on 2% agarose gels. Amplicons were purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega, UK)* (**2.10**) and sent to Source Bioscience for sequencing (**2.13**). Sequence chromatograms were visualised using SnapGene (version 5.3). VNTR types and *ompA* genotypes were assigned by the method outlined in section 2.19.2.

Table 12 Primers used for PCR amplification of VNTR loci and *ompA* in Bri038.

Primer Name	Primer sequence (5'-3')	Amplicon size[a] (bp)	Nucleotide position[a]	Reference
CT1335F*	AAAGCGTCCTCTGGAAGGG	208	737,198 - 737,216	(Labiran, 2014)
CT1335R*	CCTTCTCCTAACAACTTACGC		737,405 - 737,385	
CT1299F*	ATCGCTTAAGATTCTCGGAGG	342	291,654 - 291,674	
CT1299R*	AGGTTCTAGCTGAGCATGGG		291,995 – 291,976	
CT1291F	GCCAAGAAAAACATGCTGGT	225	195,536 – 195,555	(Pedersen *et al.*, 2008)
CT1291R	AGGATATTTCCCTCAGTTATTCG		195,760 – 195,738	
PCTM3	TCCTTGCAAGCTCTGCCTGTGGGGAATCCT	1,019	779,977 – 780,006	(Lan *et al.*, 1994)
NRI	CCGCAAGATTTTCTAGATTTC		779,008 – 778,988	

[a] According to D/UW-3/CX, Genbank accession number NC_000117.

4.4 Results

4.4.1 LGV strains of *C. trachomatis*

A total of 193 *C. trachomatis* genomes were listed in the NCBI Genome database. Of these, 133 genomes (69%) were non-LGV, hence were excluded from further analysis. The remaining 60 genomes were LGV strain genomes, however, 27 of these were laboratory-generated recombinant LGV strains. These strains were L2/434/Bu strains that were transformed. These were excluded from further analysis as we wished to genotype "wild-type" *C. trachomatis* strains and not lab-altered LGV *C. trachomatis* strains. The remaining 33 strains were LGV strains that had not been laboratory-generated. A list of the 33 LGV strains analysed in this Chapter is provided in Table 13. In total, 23 of these strains (70%) had completely assembled genome sequences, and the 10 remaining strains had in-complete genome assemblies. Whole genome shotgun data was available for these 10 strains. The ten strains are indicated in Table 13 with asterisks (*).

The *ompA* genotypes of the 33 strains analysed in this Chapter were L1 (n= 5), L2 (n= 3), L2b (n= 21), L2b/D (n= 1), and L3 (n= 1). Also included were two L1/L2 strains, L1/115 and L1/224. These strains were isolated from LGV patients in South Africa prior to 1994 (exact date unknown). *OmpA*

sequence analysis performed by Harris *et al.* (2012) determined that the sequence of variable segments 1 and 2 (VS 1 - 2) was identical to L1/440/LN, whilst the sequence of variable segments 3 and 4 (VS 3 - 4) was identical to L2/434/Bu. The L2c strain (Somboonna *et al.*, 2011) was a recombinant strain containing a chimeric L2/D genome (**Chapter 6**). The *ompA* genotype for this recombinant strain was L2.

4.4.2 Sequence Types (STs) identified

Table 14 shows the sequence types (STs) that were identified by applying the Pannekoek *et al. (*2008) MLST genotyping scheme to the 33 LGV strains. Of these 33 strains, 30 (91%) could be assigned a ST using the Pannekoek *et al.* MLST scheme. Two STs were identified – ST 44 (n= 29; 97%) and ST 11 (n= 1; 3%). ST 11 was assigned to L1/440/LN (**Table 14**).

Table 15 indicates the STs identified using the Dean *et al.* (2009) MLST genotyping system. Of the 33 LGV strains assessed, 32 (97%) could be assigned a sequence type using this system, of which 100% were assigned ST 1 (**Table 15**).

Table 16 gives the STs assigned to the 33 LGV strains using the Klint *et al*. (2007) MST genotyping system. In total, 29 strains (88%) could be assigned a sequence type. Nine STs were identified using the MST system. These included three STs that had not previously been defined, i.e. these combinations of allele types had not been reported previously. These were designated ND1 (n= 3; 10%), ND2 (n= 1; 3%), and ND3 (n= 2; 7%) (whereby ND refers to STs that were not previously defined). The ST 58 and ST 143 were exclusively assigned to L2b strains. Of the 21 L2b strains included in the analysis, 18 (86%) could be assigned a ST. Of these, 16 strains were assigned ST 58 and 2 strains were assigned ST 143 (**Table 16**).

Table 13 LGV strains of *C. trachomatis*. An asterisk (*) denotes strains with in-completely assembled genomes for which whole genome shotgun data was available.

Strain Name	*OmpA* genotype	Country (City)	Source	Collection Date	Chromosome Accession (EMBL)	Whole Genome Shotgun Accession	Genome Reference	Sequencing Method	Number of contigs
L1/115	L1/L2[a]	South Africa	LGV patient	<1994	HE601952		(Harris *et al.*, 2012)	Illumina GAII	1
L1/1322/p2	L1	South Africa	Genital ulcer	1995	HE601951		(Harris *et al.*, 2012)	Illumina GAII	5
L1/224	L1/L2[a]	South Africa	LGV patient	<1994	HE601953		(Harris *et al.*, 2012)	Illumina GAII	1
L1/440/LN	L1	USA	Lymph node	1968	HE601950		(Harris *et al.*, 2012)	Illumina GAII	2
L2/434/Bu	L2	USA	Lymph node	1968	AM884176		(Thomson *et al.*, 2007)	ABI3700	Not provided
L2/25667R	L2	USA	Rectal biopsy	<1983	HE601954		(Harris *et al.*, 2012)	Illumina GAII	1
LGV913*	L1	South Africa	Unknown	2012		CVNC00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	47
LGV1339*	L1	South Africa	Urethra	1995		CVND00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	231
LGV98*	L1	South Africa	Unknown	2012		CVNB00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	140
HPA1IMS*	L2b	UK (London)	Rectum	2005		CVOA00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	8,120*
HPA21IMS*	L2b	UK (London)	Rectum	2009		CVNY00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	419
HPA27IMS*	L2b	UK (Edinburgh)	Rectum	2005		CVOC00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	4,182*
HPA29IMS*	L2b	UK (Edinburgh)	Rectum	2004		CV0D00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	1,366*
HPA31IMS*	L2b	UK (London)	Rectum	2005		CVOE00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	584
HPA34IMS*	L2b	UK (Stoke-On-Trent)	Rectum	2008		CVPJ00000000	(Seth-Smith *et al.*, 2013)	Illumina GAII	419
L2b/795	L2b	France	Rectum	2004	HE601949		(Harris *et al.*, 2012)	Illumina GAII	1
L2b/8200/07	L2b	Sweden	Rectum	2007	HE601795		(Harris *et al.*, 2012)	Illumina GAII	1
L2b/Ams1	L2b	Netherlands	Penile ulcer	2004	HE601959		(Harris *et al.*, 2012)	Illumina GAII	2
L2b/Ams2	L2b	Netherlands	Anus	2005	HE601961		(Harris *et al.*, 2012)	Illumina GAII	7

L2b/Ams3	L2b	Netherlands	Anus	2004	HE601962		(Harris *et al.*, 2012)	Illumina GAII	8
L2b/Ams4	L2b	Netherlands	Anus	2005	HE601964		(Harris *et al.*, 2012)	Illumina GAII	8
L2b/Ams5	L2b	Netherlands	Anus	2004	HE601965		(Harris *et al.*, 2012)	Illumina GAII	4
L2b/CV204	L2b	France	Rectum	2006	HE601960		(Harris *et al.*, 2012)	Illumina GAII	11
L2b/Canada 1	L2b	Canada	Rectum	2004	HE601963		(Harris *et al.*, 2012)	Illumina GAII	4
L2b/Canada 2	L2b	Canada	Rectum	2005	HE601957		(Harris *et al.*, 2012)	Illumina GAII	3
L2b/LST	L2b	France	Rectum	2008	HE601958		(Harris *et al.*, 2012)	Illumina GAII	7
L2b/UCH-1/proctitis	L2b	UK (London)	Rectum	2006	AM884177.2		(Thomson *et al.*, 2007)	454/Roche/GS20	18
L2b/UCH-2	L2b	UK (London)	Rectum	Unknown	HE601956		(Harris *et al.*, 2012)	Illumina GAII	1
L2b/CS19/08	L2b	Portugal	Anorectal Swab	2008	CP009923		(Borges and Gomes, 2015a)	Illumina	Not provided
L2b/CS784/08	L2b	Portugal	Anorectal Swab	2008	CP009925		(Borges and Gomes, 2015a)	Illumina	Not provided
L2b/D/PT05 *	L2b/D	Portugal	Anorectal Swab	2017		CAAKND000000000	(Borges *et al.*, 2019)	Illumina	9
L2c	L2	USA	Rectum	Unknown	CP002024		(Somboonna *et al.*, 2011)	454/Roche	Not provided
L3/404/LN	L3	USA	Lymph node	1967	HE601955		(Harris *et al.*, 2012)	Illumina GAII	1

[a] *OmpA* sequence analysis of L1/L2 strains showed that variable segments 1 and 2 (VS 1 - 2) matched L1/440/LN, whilst variable segments 3 and 4 (VS 3 - 4) matched L2/434/Bu.
*fragmented assembly.

Table 14 Sequence types (STs) of 33 LGV strains determined using the Pannekoek *et al.* (2008) MLST genotyping scheme. Black boxes indicate strains that could not be assigned STs. Asterisks (*) denote strains with in-completely assembled genomes for which whole genome shotgun data was available.

Strain Name	Allele							ST
	gatA	*oppA_3*	*hflX*	*gidA*	*enoA*	*hemN*	*fumC*	
L1/115	1	3	3	3	2	2	19	44
L1/1322/p2	1	3	3	3	2	2	19	44
L1/224	1	3	3	3	2	2	19	44
L1/440/LN	1	3	3	3	2	2	2	11
L2/434/Bu	1	3	3	3	2	2	19	44
L2/25667R	1	3	3	3	2	2	19	44
LGV913*	1	3	3	3	2	2	19	44
LGV1339*	1	3	3	3	2	2	19	44
LGV98*	1	3	3	3	2	2	19	44
HPA1IMS*	1	3	3	3	2	2	19	44
HPA21IMS*	1	3	3	3	2	2	19	44
HPA27IMS*	1	3	3	3	2	2	19	44
HPA29IMS*				3	2			
HPA31IMS*	1	3	3	3	2		19	
HPA34IMS*	1	3	3	3	2	2		
L2b/795	1	3	3	3	2	2	19	44
L2b/8200/07	1	3	3	3	2	2	19	44
L2b/Ams1	1	3	3	3	2	2	19	44
L2b/Ams2	1	3	3	3	2	2	19	44
L2b/Ams3	1	3	3	3	2	2	19	44
L2b/Ams4	1	3	3	3	2	2	19	44
L2b/Ams5	1	3	3	3	2	2	19	44
L2b/CV204	1	3	3	3	2	2	19	44
L2b/Canada1	1	3	3	3	2	2	19	44
L2b/Canada2	1	3	3	3	2	2	19	44
L2b/LST	1	3	3	3	2	2	19	44
L2b/UCH-1/proctitis	1	3	3	3	2	2	19	44
L2b/UCH-2	1	3	3	3	2	2	19	44
L2b/CS19/08	1	3	3	3	2	2	19	44
L2b/CS784/08	1	3	3	3	2	2	19	44
L2b/D/PT05*	1	3	3	3	2	2	19	44
L2c	1	3	3	3	2	2	19	44
L3/404/LN	1	3	3	3	2	2	19	44

Table 15 Sequence types (STs) of 33 LGV strains determined using the Dean *et al.* (2009) MLST scheme. Black boxes indicate strains that could not be assigned STs. Asterisks (*) denote strains with incompletely assembled genomes for which whole genome shotgun data was available.

Strain Name	Allele							ST
	glyA	*mdhC*	*pdhA*	*yhbG*	*pykF*	*lysS*	*leuS*	
L1/115	1	1	3	8	1	4	11	1
L1/1322/p2	1	1	3	8	1	4	11	1
L1/224	1	1	3	8	1	4	11	1
L1/440/LN	1	1	3	8	1	4	11	1
L2/434/Bu	1	1	3	8	1	4	11	1
L2/25667R	1	1	3	8	1	4	11	1
LGV913*	1	1	3	8	1	4	11	1
LGV1339*	1	1	3	8	1	4	11	1
LGV98*	1	1	3	8	1	4	11	1
HPA1IMS*	1	1	3	8	1	4	11	1
HPA21IMS*	1	1	3	8	1	4	11	1
HPA27IMS*	1	1	3	8	1	4	11	1
HPA29IMS*		1			1	4		
HPA31IMS*	1	1	3	8	1	4	11	1
HPA34IMS*	1	1	3	8	1	4	11	1
L2b/795	1	1	3	8	1	4	11	1
L2b/8200/07	1	1	3	8	1	4	11	1
L2b/Ams1	1	1	3	8	1	4	11	1
L2b/Ams2	1	1	3	8	1	4	11	1
L2b/Ams3	1	1	3	8	1	4	11	1
L2b/Ams4	1	1	3	8	1	4	11	1
L2b/Ams5	1	1	3	8	1	4	11	1
L2b/CV204	1	1	3	8	1	4	11	1
L2b/Canada1	1	1	3	8	1	4	11	1
L2b/Canada2	1	1	3	8	1	4	11	1
L2b/LST	1	1	3	8	1	4	11	1
L2b/UCH-1/proctitis	1	1	3	8	1	4	11	1
L2b/UCH-2	1	1	3	8	1	4	11	1
L2b/CS19/08	1	1	3	8	1	4	11	1
L2b/CS784/08	1	1	3	8	1	4	11	1
L2b/D/PT05*	1	1	3	8	1	4	11	1
L2c	1	1	3	8	1	4	11	1
L3/404/LN	1	1	3	8	1	4	11	1

Table 16 Sequence types (STs) of 33 LGV strains determined using the Klint *et al.* (2007) scheme. Black boxes indicate strains that could not be assigned STs. Asterisks (*) denote strains with incompletely assembled genomes for which whole genome shotgun data was available. ND denotes a ST that has not previously been defined (i.e. this combination of allele types has not previously been reported).

Strain Name	*CT058*	*CT144*	*CT172*	*hctB*	*pbpB*	ST
L1/115	11	17	10	18	28	ND1
L1/1322/p2	13	4	10	18	28	ND2
L1/224	11	17	10	18	28	ND1
L1/440/LN	11	4	6	18	28	49
L2/434/Bu	13	19	6	18	28	141
L2/25667R	37	19	6	18	28	144
LGV913*	13	4	6	18	28	ND3
LGV1339*	13	4	6	18	28	ND3
LGV98*	11	17	10	18	28	ND1
HPA1IMS*	13	17	13			
HPA21IMS*	13	17	13	27	28	58
HPA27IMS*	13	17	13	44	28	143
HPA29IMS*			13			
HPA31IMS*	13	17	13	44		
HPA34IMS*	13	17	13	44	28	143
L2b/795	13	17	13	27	28	58
L2b/8200/07	13	17	13	27	28	58
L2b/Ams1	13	17	13	27	28	58
L2b/Ams2	13	17	13	27	28	58
L2b/Ams3	13	17	13	27	28	58
L2b/Ams4	13	17	13	27	28	58
L2b/Ams5	13	17	13	27	28	58
L2b/CV204	13	17	13	27	28	58
L2b/Canada1	13	17	13	27	28	58
L2b/Canada2	13	17	13	27	28	58
L2b/LST	13	17	13	27	28	58
L2b/UCH-1/proctitis	13	17	13	27	28	58
L2b/UCH-2	13	17	13	27	28	58
L2b/CS19/08	13	17	13	27	28	58
L2b/CS784/08	13	17	13	27	28	58
L2b/D/PT05*	13	17	13		28	
L2c	13	19	6	18	28	141
L3/404/LN	11	4	10	19	31	51

4.4.3 MLVA-*ompA* genotypes identified

Table 17 indicates the MLVA-*ompA* genotypes identified using the Pedersen *et al.* (2008) MLVA-*ompA* genotyping system. Of the 33 LGV strains assessed, 100% could be assigned MLVA-*ompA* genotypes. In total, eight distinct MLVA-*ompA* genotypes were identified. The 21 L2b strains were resolved into 1 MLVA-*ompA* genotype using the system – 1.9.2b - L2b. Table 18 shows the VNTR sequence variants identified. Application of MLVA-*ompA* identified one new CT1291 variant that had not previously been seen in the published literature. This VNTR variant was assigned CT1291 type 4b (AAAATAGTCTA-**10C**-TATTG) (**Table 18**).

4.4.4 Discriminatory Power and Typeability of the Genotyping Systems for *C. trachomatis*

Table 19 provides a summary of the STs and MLVA-*ompA* genotypes identified by applying the four genotyping systems – MST (Klint *et al.*, 2007), MLST (Pannekoek *et al.*, 2008), MLST (Dean *et al.*, 2009), and MLVA-*ompA* (Pedersen *et al.*, 2008) – to the 33 LGV strains of *C. trachomatis* provided in Table 13.

The discriminatory power index (D) was calculated for each system: $D_{MLVA\text{-}ompA} = 0.59$; $D_{MST} = 0.75$; D_{MLST} (Pannekoek *et al*, 2008.) = 0.22; and D_{MLST} (Dean *et al.*, 2009) = 0.06 (**Table 19**).

The typeability of each genotyping system was calculated: MLVA-*ompA* (Pedersen *et al.*, 2008) = 100%, MST (Klint *et al.*, 2007) = 88%, MLST (Pannekoek *et al.*, 2008) = 91%, and MLST (Dean *et al.*, 2009) = 97% (**Table 19**).

Table 17 MLVA-*ompA* genotypes of 33 LGV strains determined using the Pedersen *et al.* (2008) MLVA-*ompA* scheme. Asterisks (*) denote strains with in-completely assembled genomes for which whole genome shotgun data was available. † indicates a VNTR type that has not previously been described in the literature by Wang *et al.* (2011), Satoh *et al.* (2014), Qin *et al.* (2016), or Labiran *et al.* (2016, 2017).

Strain Name	CT1335	CT1299	CT1291	*ompA*	MLVA-*ompA* genotype
L1/115	5	9	8	L1	5.9.8 - L1
L1/1322/p2	5	9	8	L1	5.9.8 - L1
L1/224	5	9	8	L1	5.9.8 - L1
L1/440/LN	5	9	3b	L1	5.9.3b - L1
L2/434/Bu	5	9	3b	L2	5.9.3b - L2
L2/25667R	5	9	3b	L2	5.9.3b - L2
LGV913*	5	9	4b†	L1	5.9.4b† - L1
LGV1339*	5	9	2b	L1	5.9.2b - L1
LGV98*	5	9	8	L1	5.9.8 - L1
HPA1IMS*	1	9	2b	L2b	1.9.2b - L2b
HPA21IMS*	1	9	2b	L2b	1.9.2b - L2b
HPA27IMS*	1	9	2b	L2b	1.9.2b - L2b
HPA29IMS*	1	9	2b	L2b	1.9.2b - L2b
HPA31IMS*	1	9	2b	L2b	1.9.2b - L2b
HPA34IMS*	1	9	2b	L2b	1.9.2b - L2b
L2b/795	1	9	2b	L2b	1.9.2b - L2b
L2b/8200/07	1	9	2b	L2b	1.9.2b - L2b
L2b/Ams1	1	9	2b	L2b	1.9.2b - L2b
L2b/Ams2	1	9	2b	L2b	1.9.2b - L2b
L2b/Ams3	1	9	2b	L2b	1.9.2b - L2b
L2b/Ams4	1	9	2b	L2b	1.9.2b - L2b
L2b/Ams5	1	9	2b	L2b	1.9.2b - L2b
L2b/CV204	1	9	2b	L2b	1.9.2b - L2b
L2b/Canada1	1	9	2b	L2b	1.9.2b - L2b
L2b/Canada2	1	9	2b	L2b	1.9.2b - L2b
L2b/LST	1	9	2b	L2b	1.9.2b - L2b
L2b/UCH-1/proctitis	1	9	2b	L2b	1.9.2b - L2b
L2b/UCH-2	1	9	2b	L2b	1.9.2b - L2b
L2b/CS19/08	1	9	2b	L2b	1.9.2b - L2b
L2b/CS784/08	1	9	2b	L2b	1.9.2b - L2b
L2b/D/PT05*	1	9	2b	L2b/D	1.9.2b - L2b/D
L2c	5	9	3b	L2	5.9.3b - L2
L3/404/LN	5	9	8	L3	5.9.8 - L3

Table 18 Variable Number Tandem Repeat (VNTR) sequence variants identified using the Pedersen *et al.* (2008) MLVA-*ompA* scheme. Asterisks (*) denote strains with incompletely assembled genomes for which whole genome shotgun data was available. † indicates a VNTR type that has not previously been described by Wang *et al.* (2011), Satoh *et al.* (2014), Qin *et al.* (2016), or Labiran *et al.* (2016, 2017).

Strain Name	CT1335	CT1299	CT1291
L1/115	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**7C**-TATTG
L1/1322/p2	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**7C**-TATTG
L1/224	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**7C**-TATTG
L1/440/LN	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**9C**-TATTG
L2/434/Bu	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**9C**-TATTG
L2/25667R	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**9C**-TATTG
LGV913*	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**10C**-TATTG †
LGV1339*	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
LGV98*	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**7C**-TATTG
HPA1IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
HPA21IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
HPA27IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
HPA29IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
HPA31IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
HPA34IMS*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/795	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/8200/07	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Ams1	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Ams2	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Ams3	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Ams4	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Ams5	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/CV204	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Canada1	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/Canada2	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/LST	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/UCH-1/proctitis	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/UCH-2	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/CS19/08	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/CS784/08	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2b/D/PT05*	GAAAAAG-**9T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**8C**-TATTG
L2c	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**9C**-TATTG
L3/404/LN	GAAAAAG-**11T8A**-GCTTTTGT	TTTTTATTCT-3C2T-**6C**-ATCAAA	AAAATAGTCTA-**7C**-TATTG

Table 19 Summary of Sequence Types (STs) and MLVA-*ompA* genotypes assigned to 33 LGV *C. trachomatis* strains. Discriminatory index (D) and typeability of each genotyping scheme is given. Black boxes indicate strains that could not be assigned STs or MLVA-*ompA* genotypes. ND=STs that have not previously been defined; †VNTR types that have not previously been described; *strains with incompletely assembled genomes for which whole genome shotgun data was available.

Strain Name	MLVA-*ompA* genotype (Pedersen *et al.*, 2008)	MLST-7 ST (Pannekoek *et al.*, 2008)	MLST-7[a] ST (Dean *et al.*, 2009)	MST ST (Klint et al., 2007)
L1/115	5.9.8 - L1	44	1	ND1
L1/1322/p2	5.9.8 - L1	44	1	ND2
L1/224	5.9.8 - L1	44	1	ND1
L1/440/LN	5.9.3b - L1	11	1	49
L2/434/Bu	5.9.3b - L2	44	1	141
L2/25667R	5.9.3b - L2	44	1	144
LGV913*	5.9.4b† - L1	44	1	ND3
LGV1339*	5.9.2b - L1	44	1	ND3
LGV98*	5.9.8 - L1	44	1	ND1
HPA1IMS*	1.9.2b - L2b	44	1	
HPA21IMS*	1.9.2b - L2b	44	1	58
HPA27IMS*	1.9.2b - L2b	44	1	143
HPA29IMS*	1.9.2b - L2b			
HPA31IMS*	1.9.2b - L2b		1	
HPA34IMS*	1.9.2b - L2b		1	143
L2b/795	1.9.2b - L2b	44	1	58
L2b/8200/07	1.9.2b - L2b	44	1	58
L2b/Ams1	1.9.2b - L2b	44	1	58
L2b/Ams2	1.9.2b - L2b	44	1	58
L2b/Ams3	1.9.2b - L2b	44	1	58
L2b/Ams4	1.9.2b - L2b	44	1	58
L2b/Ams5	1.9.2b - L2b	44	1	58
L2b/CV204	1.9.2b - L2b	44	1	58
L2b/Canada1	1.9.2b - L2b	44	1	58
L2b/Canada2	1.9.2b - L2b	44	1	58
L2b/LST	1.9.2b - L2b	44	1	58
L2b/UCH-1/proctitis	1.9.2b - L2b	44	1	58
L2b/UCH-2	1.9.2b - L2b	44	1	58
L2b/CS19/08	1.9.2b - L2b	44	1	58
L2b/CS784/08	1.9.2b - L2b	44	1	58
L2b/D/PT05*	1.9.2b - L2b/D	44	1	
L2c	5.9.3b - L2	44	1	141
L3/404/LN	5.9.8 - L3	44	1	51
Number of variants identified	8	2	1	9
Typeability	33/33 (100%)	30/33 (91%)	32/33 (97%)	29/33 (88%)
Discriminatory Index (D)	0.59	0.22	0.06	0.75

4.4.1 A Field-Test of the MLVA-*ompA* genotyping system

A field-test of the MLVA-*ompA* genotyping system was conducted using genomic DNA extracted from a *C. trachomatis* LGV isolate, Bri038 (**Chapter 6**). This isolate had previously been MLVA-*ompA* genotyped by Labiran *et al.* (2016) and assigned the genotype 1.9.2 – L2b. The aim of the field-test was to assess the practicability of the MLVA-*ompA* system and the reproducibility of the MLVA-*ompA* genotyping system (i.e. the ability to assign the same genotype upon repeat testing).

The three VNTR loci (CT1335, CT1299, and CT1291) and *ompA* were amplified by PCR using the primers listed in Table 12. The PCR products for the four markers were visualised on 2% agarose gels (**Figure 17**). For each marker, a single PCR product was visible in the lanes loaded with Bri038, and the lanes loaded with the positive control. The PCR products for each marker were subsequently purified (**2.10**) and sequenced (**2.13**).

The sequence chromatograms of the three VNTR loci (CT1335, CT1299, and CT1291) for Bri038 were visualised (**Figure 18**). The VNTR types for Bri038 were CT1335 type 1 (GAAAAAG-**9T8A**-GCTTTTGT), CT1299 type 9 (TTTTTATTCT-3C2T-**6C**-ATCAAA), and CT1291 type 2b (AAAATAGTCTA-**8C**-TATTG), corresponding to a MLVA type code of 1.9.2b. Labiran *et al.* (2016) recorded the VNTR type of Bri038 as 1.9.2 and not 1.9.2b. To check that the VNTR sequences that were obtained in this Chapter were identical to those obtained for Bri038 by Labiran *et al.* (2016); the sequence chromatograms obtained by Labiran were visualised. The VNTR type at CT1291 in the chromatogram obtained by Labiran was also type 2b, however Labiran recorded the CT1291 VNTR type as type 2 mistakenly.

The *ompA* sequence of Bri038 was aligned against that of L2b/UCH-1/proctitis (Genbank accession no. AM884177) and L2/434/Bu (Genbank accession no. AM884176) to determine the *ompA* genotype (**Appendix 3**). Also included in the alignment for comparison was the *ompA* sequence of Bri038 obtained by Labiran *et al.* (2016). Appendix 3 shows that the *ompA* sequence of Bri038 obtained in this Chapter is identical to the *ompA* sequence obtained for this isolate by Labiran *et al*. (**Appendix 3**). The Bri038 *ompA* sequence contains a SNP at nucleotide position 485 (A485G). This SNP is present in L2b (and not L2) *C. trachomatis* strains. A second SNP, at nucleotide position 517 (C517A) was also present. The MLVA-*ompA* genotype of Bri038 was therefore determined to be 1.9.2b - L2b. This matched the MLVA-*ompA* genotype obtained by Labiran *et al.* for this isolate.

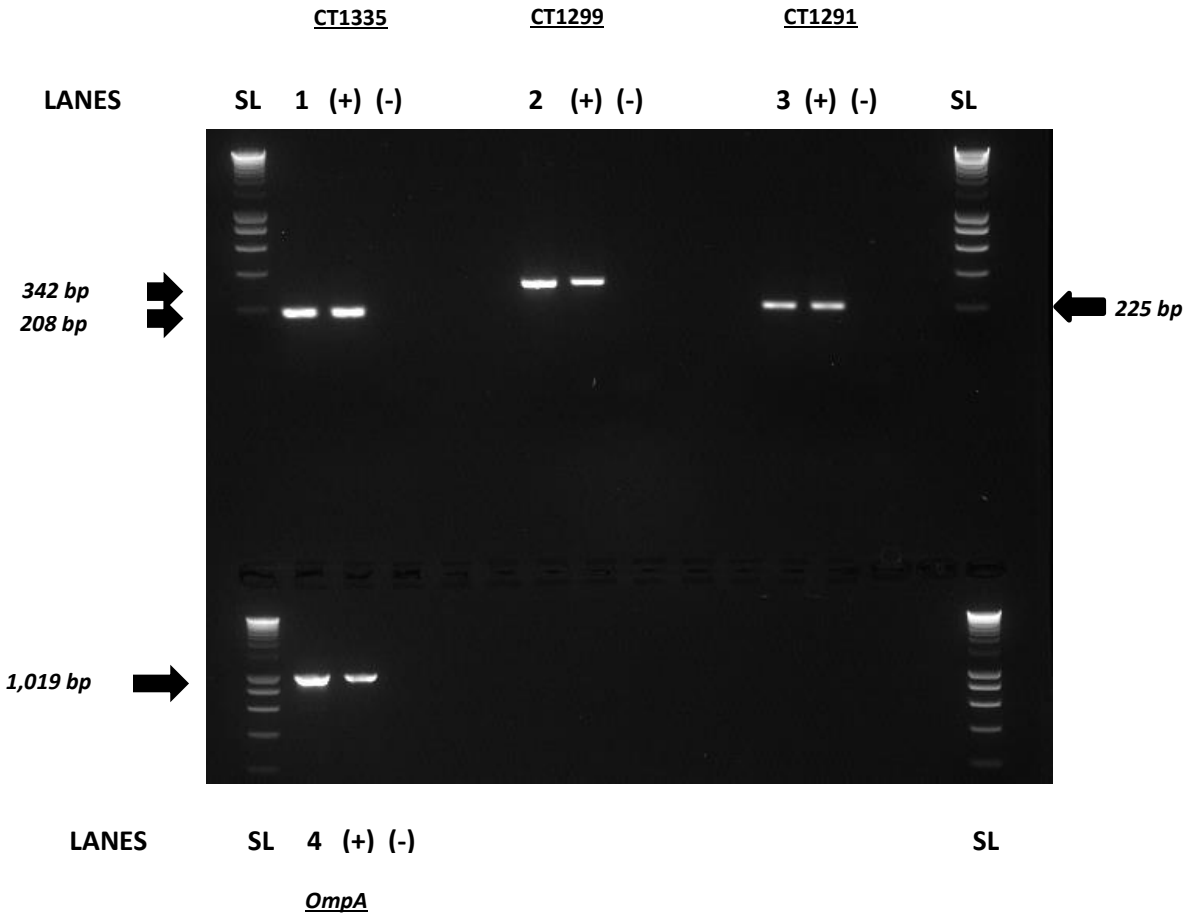


Figure 17 A 2% agarose gel showing the PCR products after amplification of three VNTR regions (CT1335, CT1299 and CT1291) and *ompA* from Bri038. The lane labelled "1" shows the products from amplification of CT1335 in Bri038; the lane labelled "2" shows the products from amplification of CT1299 in Bri038; the lane labelled "3" shows the products of amplification of CT1291 in Bri038; and the lane labelled "4" shows the products of amplification of *ompA* in Bri038. SL represents the DNA ladder Smartladder. (+) represents the positive control, SotonK1. (-) represents the no-template negative control (NTC).

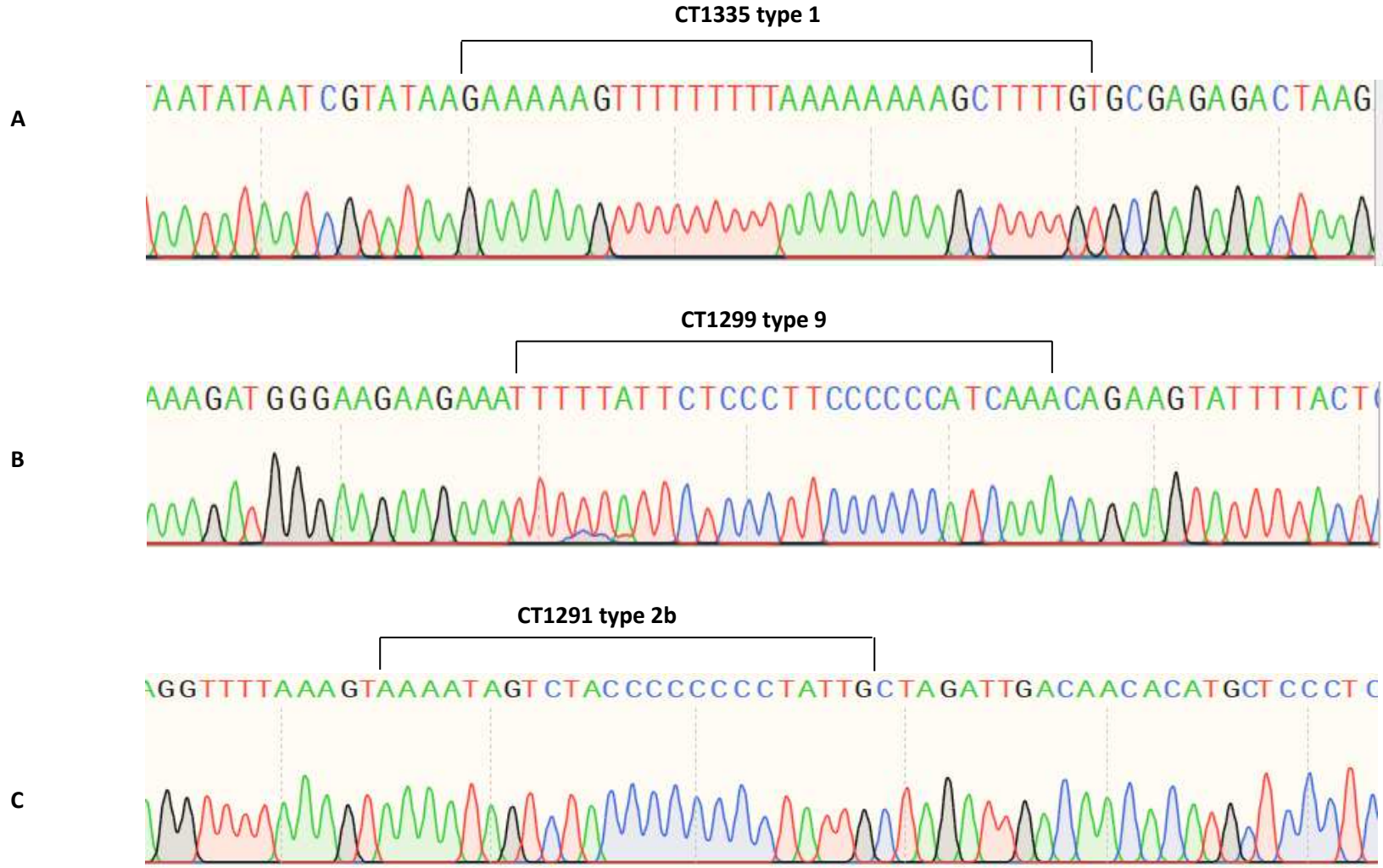


Figure 18 Chromatograms showing the three VNTR regions (CT1335, CT1299, and CT1291) in Brighton isolate Bri038. The sequence chromatograms of (A) CT1335, (B) CT1299, and (C) CT1291 are shown. The corresponding VNTR type is indicated in bold above each chromatogram.

4.5 Discussion

The aim of this Chapter was to evaluate four high-resolution multi-locus genotyping systems for *C. trachomatis* (Klint *et al.*, 2007; Pannekoek *et al.*, 2008; Pedersen *et al.*, 2008; Dean *et al.*, 2009) in order to determine which system would be most suitable to apply in a genotyping survey of LGV strains collected from MSM in London (**Chapter 5**). To achieve this aim, we took an in silico data approach and obtained whole genome sequence data for all LGV strains deposited in NCBI Genome. We excluded any laboratory-generated LGV strains from this analysis, opting to only genotype "wild-type" LGV strains (i.e. non-laboratory modified). The reason for this was to ensure that our results were as close as possible to those that could be obtained from real-world clinical LGV specimens. We identified a total of 33 LGV strains using the NCBI Genome database that matched this criterion. The sequence types (STs) and MLVA-*ompA* genotypes were determined for each of these 33 strains. We assessed the discriminatory power of each genotyping system (i.e. the ability of the genotyping system to discriminate between un-related strains) by calculating Simpson's Index of Diversity (van Belkum *et al.*, 2007), and calculated the typeability for each genotyping system.

Our results showed that the MLST schemes developed by Pannekoek *et al.* (2008) and Dean *et al.* (2009) lacked resolution and could not sufficiently distinguish between the LGV strains analysed. The Dean *et al.* MLST scheme resolved the LGV strains into one single sequence type (ST), ST 1. The Pannekoek *et al.* MLST scheme resolved the LGV strains into 2 sequence types, ST 11 and ST 44. However, ST 44 was assigned to 97% of the LGV strains analysed using this typing system, and ST 11 was assigned to one LGV strain only (L1/440/LN). The discriminatory index of both systems was correspondingly low. The discriminatory power for the Pannekoek *et al.* (2008) MLST scheme was 0.22, and the discriminatory power of the Dean *et al.* (2009) MLST scheme was 0.06. Given that typing systems with Index values above 0.95 are considered to have more or less "ideal" resolution, the Pannekoek *et al.* (2008) and Dean *et al.* (2009) MLST systems can be seen to have very low resolution. However, it should be noted that these two MLST schemes are based on housekeeping genes that are stable and not under immune selection. In addition, these schemes were developed to describe intra-specific variation and population structure rather than to operate as ultimate strain typing schemes (Urwin and Maiden, 2003). The low resolution provided by these two MLST schemes makes them unsuitable candidates for application in the LGV genotyping survey in London (**Chapter 5**).

The Klint *et al.* MST scheme (Klint *et al.*, 2007) is based on five hypervariable genetic loci, including two non-housekeeping genes, *pbpB* and *hctB*. These two genes are subject to selection pressure

which introduces sequence variation. As such, applications of MST include short-term outbreak investigations and local epidemiological studies. In this Chapter, the Klint *et al.* MST scheme was able to discriminate the LGV strains analysed into nine sequence types, including three STs that had not previously been reported. The discriminatory index of this system was the highest of all four systems evaluated in this Chapter (D_{MST} = 0.75). The MST scheme was the only system evaluated that was able to resolve L2b strains into more than one ST (ST 58 and ST 143). However, the typeability of the system was the lowest of all four genotyping schemes evaluated, and STs could not be assigned to four strains (12%). The lower typeability obtained from MST is likely due to the fact that some of the genomes included in the in silico analyses were highly fragmented with a large number of contigs (HPA1IMS – HPA34IMS had contig numbers ranging from 419 to 8,120). A highly fragmented genome produces difficulties when analysing sequence variation and would explain why these strains were unable to be assigned a type at larger gene loci (*pbpB* and *hctB*) by MST.

Our results indicate that the MLVA-*ompA* system had the highest typeability (100%) of all four systems, with MLVA-*ompA* genotypes assigned to all 33 LGV strains. The system was able to resolve LGV strains into eight MLVA-*ompA* genotypes, and identified one new VNTR sequence variant that had not previously been defined in the published literature (Wang *et al.*, 2011; Satoh *et al.*, 2014; Labiran *et al.*, 2016, 2017; Qin *et al.*, 2016): CT1291 type 4b (AAAATAGTCTA-**10C**-TATTG) (**Table 17**). However, the MLVA-*ompA* system was able to resolve the 18 L2b strains into one MLVA-*ompA* genotype only, 1.9.2b - L2b. The discriminatory index of the system was 0.59, which is lower than the Klint *et al.* MST system (D_{MST} = 0.76), but also lower than the discriminatory index values for MLVA-*ompA* that have been previously reported in other studies (between 0.94 and 0.99) (Pedersen *et al.*, 2008; Ikryannikova *et al.*, 2010; Labiran *et al.*, 2017). It is possible that the strains included in this Chapter were not a particularly diverse population, and as a result, the genotypic diversity in LGV strains is not being accurately reflected. A MLVA-*ompA* genotyping study by Labiran *et al.* (2016) identified a high level of genotypic diversity amongst L2b isolates, assigning seven distinct MLVA-*ompA* genotypes to nine L2b isolates. This contrasts with the single MLVA-*ompA* genotype identified amongst L2b strains in this Chapter.

We have shown that both MST and MLVA-*ompA* were able to resolve the LGV strains into eight and nine MLVA-*ompA* genotypes and STs, respectively. Hence, both are excellent candidates for application in the genotyping survey of LGV clinical specimens collected in London (**Chapter 5**). However, the practicability of both schemes must be considered when selecting the most suitable system for a large-scale genotyping survey. A limitation of MST is that two of the regions (CT058 and *pbpB*) are over 1,500 bp in length and require multiple internal primers for sequencing. The original Klint *et al.* MST system has been adapted to a nested PCR (Bom *et al.*, 2011), however the adapted method is also problematic. Nested PCR is highly sensitive but highly susceptible to

contamination as it involves more sample manipulation (da Silva *et al.*, 2013). In contrast, MLVA-*ompA* targets short (≤225 bp) VNTR sequences in addition to the *ompA* gene (1,019 bp), and each can be amplified by a single set of primers. This is highly advantageous when the number of clinical specimens to be processed in the genotyping survey (**Chapter 5**) is considered.

It was decided to conduct a field-test for the MLVA-*ompA* genotyping system using genomic DNA extracted from a rectal isolates, Bri038, collected in Brighton (Labiran *et al.*, 2016). This isolate had previously been MLVA-*ompA* genotyped by Labiran *et al.* (2016). It was decided to repeat the MLVA-*ompA* analysis conducted on the isolates to assess the reproducibility of the MLVA-*ompA* genotyping system (i.e. the ability to assign the same MLVA-*ompA* genotype to an isolate upon repeated testing). We identified that the MLVA-*ompA* genotype of Bri038 was 1.9.2b - L2b. The *ompA* sequence obtained from Bri038 in this Chapter, and that obtained from Bri038 by Labiran *et al.* (2016) were aligned. The sequences were found to be identical. In terms of the practicability of MLVA-*ompA*, the PCR set-up was not time-consuming, and the assignation of VNTR types and *ompA* genotypes was straight-forward. We can now conclude that the MLVA-*ompA* genotyping system will be applied to genotype LGV clinical specimens in Chapter 5.

CHAPTER 5 Diversity of Lymphogranuloma venereum (LGV) Strains in High-Risk Groups in London

Part of the work described in this Chapter has been published in Manning *et al.* (2021) (**Appendix 4**).

5.1 Introduction

LGV is diagnosed almost exclusively amongst men who have sex with men (MSM) (**1.8.3** and **1.10**), and LGV epidemics in MSM have been reported in Europe, Australia, North America, and Canada (Martin-Iguacel *et al.*, 2010). LGV infections in MSM typically manifest as proctitis, however asymptomatic infections have been reported in approximately 27% - 43% of cases (Spaargaren, Fennema, *et al.*, 2005; Saxon *et al.*, 2016) (**1.10.2**). LGV testing is provided for asymptomatic HIV-positive MSM with a *C. trachomatis* diagnosis (Nwokolo *et al.*, 2016), however not all asymptomatic LGV cases will be detected and treated, and these un-diagnosed cases can form a reservoir of *C. trachomatis* infection (Saxon *et al.*, 2016) that can spread un-detected through a high-risk population (**1.8.3**).

The UK has the largest documented LGV outbreak amongst MSM in Europe, and in 2019, the number of LGV cases in MSM in England (n= 1,076) was the highest recorded since UK-wide LGV surveillance began (Public Health England, 2020d). Asymptomatic carriage of LGV amongst MSM may partially explain the rise in cases in England (Public Health England, 2020d), however increased screening in MSM may also have contributed to a reported increase in case numbers (Public Health England, 2020d). The diagnostic rates of LGV are highest amongst MSM living with HIV, and it has been suggested that this may be due to HIV seroadaptive behaviours including serosorting (**1.8.3**) (Hughes *et al.*, 2013; Rönn *et al.*, 2014; Public Health England, 2020d). Although serosorting can offer some benefit in terms of reduced risk of HIV transmission, it provides no protection against other STIs including LGV (Grov *et al.*, 2015).

Whilst the rate of LGV is higher amongst HIV-positive MSM, a report published by Public Health England has shown that the diagnostic rate of LGV amongst HIV-negative MSM has steadily increased in recent years, with a 95% increase between 2018 and 2019 (Public Health England, 2020d). It has been suggested that the increase in cases within HIV-negative MSM could be associated with the increasing availability of pre-exposure prophylaxis (PrEP). PrEP use has been linked to an increase in sexual risk-taking including condom-less sex and number of sexual partners (**1.8.3** and **1.10.5.2**) (Gafos *et al.*, 2019; Traeger *et al.*, 2019). In England, LGV cases are concentrated

in London (77% of all cases in 2019 were in London), and cases in London have risen by 34% between 2018 and 2019 (Public Health England, 2020d).

Amongst MSM diagnosed with LGV, *ompA* genotype L2 was the dominant circulating genotype until 2003, when *ompA* genotype L2b was identified as the cause of an LGV outbreak in a sexual network of MSM with proctitis in Rotterdam (Götz *et al.*, 2004; Spaargaren, Fennema, *et al.*, 2005) (**1.10**). A single SNP (A485G) causing an amino acid substitution (N162S) in variable segment 2 of the *ompA* gene (**1.11**) distinguishes *ompA* of L2b strains from L2 *C. trachomatis* strains (Seth-Smith *et al.*, 2021). The L2b variant has since been implicated in LGV outbreaks worldwide (Nieuwenhuis *et al.*, 2004; Spaargaren *et al.*, 2005; Ward *et al.*, 2007; Rodriguez-Dominguez *et al.*, 2015; Isaksson *et al.*, 2017; Cole *et al.*, 2020b). It was retrospectively identified that *C. trachomatis* isolates with *ompA* genotype L2b were circulating in the US between 1979 and 1985 (Spaargaren *et al.*, 2005; Christerson *et al.*, 2010) (**1.10.1**). It has been suggested that the European LGV outbreaks may have originated in North America, and are the result of clonal L2b expansion facilitated by global travel and involvement of MSM in international sexual networks (Seth-Smith *et al.*, 2013; Borges *et al.*, 2019).

Over the past decade, a resurgence in the L2 *ompA* genotype has been reported in Austria and Croatia (Cole *et al.*, 2020a), Spain (Rodríguez-Domínguez *et al.*, 2014), France (Peuchant *et al.*, 2016) and Sweden (Isaksson *et al.*, 2017); and in these countries, strains with *ompA* genotype L2 now dominate. A retrospective study of LGV *C. trachomatis* strains collected in France between 2010 and 2015 (Peuchant *et al.*, 2016) found that the proportion of LGV cases caused by L2b declined after 2012, whilst the proportion caused by *ompA* genotype L2 increased from 2012 onwards (**Chapter 6**). In contrast, *ompA* genotyping of LGV samples from the UK (Cole *et al.*, 2020b) in 2020 showed that the L2b, and not L2, *ompA* genotype predominated. However, this study only collected a small number of LGV-positive clinical specimens (n = 22) for *ompA* genotyping. To more accurately assess sequence diversity within LGV strains circulating in the UK, a larger-scale genotyping survey of LGV strains is required.

Novel L2 and L2b *ompA* variants have been detected across Europe (Stary *et al.*, 2008; Cole *et al.*, 2020b; Marangoni *et al.*, 2021; Seth-Smith *et al.*, 2021), including the L2bV1 variant reported in Austria (Cole *et al.*, 2020b); L2bV5, identified in four samples from the UK (Cole *et al.*, 2020b); and L2bV6, identified in one sample from the UK in 2020 (Cole *et al.*, 2020b). High levels of recombination within *ompA* have been detected (Brunelle and Sensabaugh, 2006; Borges and Gomes, 2015b). Recombination is a driver of *C. trachomatis* evolution (Joseph *et al.*, 2012; Seth-Smith *et al.*, 2013; Hadfield *et al.*, 2017). One recombination event between the *ompA* gene of an *ompA* genotype D strain and the L2b genomic backbone has resulted in a hybrid L2b/D *C.*

trachomatis strain identified in Portugal in 2017 (Borges *et al.*, 2019). The transmission and pathogenic capabilities of this hybrid L2b/D strain are still unknown (Borges *et al.*, 2019).

Genotyping systems that explore polymorphisms at multiple loci can provide a more accurate depiction of genetic variation within *C. trachomatis* than typing systems based on a single locus (such as *ompA* genotyping) (van Belkum *et al.*, 2007) (**1.11**). High-resolution genotyping systems can be applied to determine the distribution of genotypes circulating within populations (Gravningen *et al.*, 2012) and to track the transmission of circulating strains within sexual networks (Klint *et al.*, 2007; Labiran *et al.*, 2016). The MLVA-*ompA* genotyping system (Pedersen *et al.*, 2008) has been successfully applied to genotype *C. trachomatis* strains globally (Wang *et al.*, 2011; Peuchant *et al.*, 2012; Satoh *et al.*, 2014; Herrmann *et al.*, 2015; Labiran *et al.*, 2016, 2017), and in 2016, the system was applied to clinical specimens collected from the rectum of MSM in Brighton (Labiran *et al.*, 2016). The study detected a high level of genotypic diversity within LGV strains: nine distinct LGV MLVA-*ompA* genotypes were identified within eleven LGV-positive specimens, with seven distinct LGV MLVA-*ompA* genotypes detected within specimens assigned L2b. No large-scale genotyping survey of LGV strains using the MLVA-*ompA* genotyping system has been conducted in the United Kingdom, hence it is unknown whether the high level of genotypic diversity within LGV strains in Brighton is also present within LGV strains circulating in other UK cities.

The hypothesis for this Chapter was that a high level of genotypic diversity may be found within LGV strains in London, given that London has the highest incidence of LGV in the UK. It was decided to conduct a preliminary genotyping survey of LGV strains in London. The data collected on genotypic diversity within the London population could later be used to determine whether a larger, multi-city genotyping survey of LGV strains should be carried out in the future, to assess genotypic diversity in LGV strains in other UK cities. In a previous Chapter, the MLVA-*ompA* genotyping system was shown to be most well-suited to genotype LGV strains (**Chapter 4**). Hence, it was decided to apply MLVA-*ompA* to genotype LGV-positive clinical specimens collected from London.

5.2 Aims and Objectives

The aim of this Chapter was to assess genotypic diversity within LGV strains in London, by determining the distribution of LGV MLVA-*ompA* genotypes in LGV-positive clinical specimens from eight London genitourinary medicine (GUM) clinics.

To achieve this aim, the objectives were as follows:

- To gain ethical approval to collect and genotype LGV-positive clinical specimens from eight London-based GUM clinics
- To apply the MLVA-*ompA* genotyping system (Pedersen *et al.*, 2008) and assign MLVA-*ompA* genotypes to LGV-positive specimens collected from eight London GUM clinics
- To determine the typeability of the MLVA-*ompA* genotyping system, i.e. - the percentage of samples assigned a MLVA-*ompA* genotype out of the total number of samples analysed.
- To determine the discriminatory power of the MLVA-*ompA* genotyping system, using Hunter-Gaston's modification of Simpson's Index of Diversity (Hunter and Gaston, 1988) (**2.20.1**).

5.3 Study Design

The study was a high-resolution genotyping survey of LGV strains utilising the MLVA-*ompA* genotyping system (Pedersen *et al.*, 2008). The study was designed to assess genotypic diversity within LGV strains in London. London was selected as the location for the study as it reports the highest incidence of LGV in the United Kingdom. Hence, basing the study in London would allow us to collect and genotype a large number of clinical specimens from across a wide range of sexual networks.

To determine the London Genitourinary Medicine (GUM) clinics to approach for involvement in the study, we consulted Dr John Saunders, a Medical Consultant in the Blood Safety, Hepatitis, STI and HIV Division at Public Health England. To keep logistics as simple as possible, it was decided to collect clinical specimens from clinics reporting the highest number of diagnoses of LGV. This would maximise the number of clinical specimens that could be collected from each clinic. Dr Saunders provided a list of GUM clinics across England that reported the highest number of LGV diagnoses in MSM between 1st January 2017 and 31st December 2017 (the list was generated in January 2018, hence this period covered the preceding 12 months). The list is shown in Table 20. The list was generated using data on diagnoses available on Genitourinary Medicine Clinic Activity Dataset (GUMCAD). GUMCAD is the national STI surveillance system and collects data on STI tests and diagnoses from all sexual health services in England (**1.10.5.1**). It should be noted that Table 20 includes only diagnoses of LGV within MSM, and excludes diagnoses made in females and heterosexual males during the time period. This was due to their infrequency – there were 18 diagnoses made in females, and 20 diagnoses in heterosexual males, across all GUM clinics in England in 2017.

Table 20 GUM[1] clinics in England reporting the highest number of LGV diagnoses in MSM[2 †] between 01/01/2017 and 31/12/2017. Data was collected from GUMCAD[3] by Dr John Saunders (Personal Communication).

GUM[1] Clinic	**Number of LGV Diagnoses in MSM[2 †] in 2017**
56 Dean Street, London	150
Burrell Street Sexual Health Centre, London	50
Mortimer Market Centre, London	31
John Hunter Clinic, London	25
The Hathersage Centre, Manchester	23
Claude Nichol Centre, Brighton	11
Total number of diagnoses made in MSM in 2017	290

[1] GUM = Genitourinary Medicine Clinic; [2] MSM = Men who have sex with men; [†] MSM includes men who reported being gay or bisexual; [3] GUMCAD = Genitourinary Medicine Clinic Activity Dataset.

Table 20 identified that four of the six GUM clinics reporting the highest number of LGV diagnoses amongst MSM in 2017 were based in London - 56 Dean Street, Burrell Street Sexual Health Centre, Mortimer Market Centre, and John Hunter Clinic (**Table 20**). We contacted Consultant Physicians at each of these four GUM clinics to enquire about participation in the genotyping study and received responses from Dr Gary Whitlock (56 Dean Street) and Dr Achyuta Nori (Burrell Street Sexual Health Centre).

At GUM clinics, clinical specimens ranging from rectal swabs, throat swabs, urine or "pooled 3-in-1" specimens (a rectal swab and pharyngeal swab in urine, from a single patient) are obtained from patients by clinicians or nurses. These clinical specimens are subsequently sent to pathology providers for *C. trachomatis* testing.

Clinical specimens from 56 Dean Street are tested for *C. trachomatis* by North West London Pathology (NWLP). NWLP is an NHS pathology partnership between Imperial College Healthcare NHS Trust, Chelsea & Westminster NHS Foundation Trust, and the Hillingdon Hospitals NHS Foundation Trust. It is hosted by Imperial College Healthcare NHS Trust and is based at Charing Cross Hospital. NWLP also provides *C. trachomatis* testing for clinical specimens collected from

patients attending John Hunter Clinic, 10 Hammersmith Broadway, and Jefferiss Wing. At NWLP, clinical specimens are tested for *C. trachomatis* using the BD ProbeTec™ ET *C. trachomatis*/*N. gonorrhoeae* Amplified DNA Assay (*Becton, Dickinson and Company, USA*). This is a nucleic acid amplification test (NAAT) that amplifies and detects *C. trachomatis* DNA in clinical specimens using amplification primers and a fluorescent labelled detector probe. At NWLP, all rectal NAAT-positive specimens are subsequently tested for LGV. Nucleic acids are isolated from specimens using the MagNA Pure Compact Nucleic Acid Isolation Kit I (*Roche Life Science, UK*), and LGV confirmation testing (**1.10.3**) is undertaken using a triplex LGV PCR assay developed by the Sexually Transmitted Bacteria Reference Laboratory (STBRL) at Public Health England (PHE) (**Appendix 5**) (Morré *et al.*, 2005; Alexander, Martin and Ison, 2007; Chen *et al.*, 2007). This assay targets the *pmpH* gene located on the *C. trachomatis* chromosome; in addition to an 88 bp region of the *C. trachomatis* cryptic plasmid; and the human *RNase P* gene as an internal control. The assay distinguishes LGV from non-LGV *C. trachomatis* strains, as it targets a unique 36 bp deletion in the polymorphic membrane protein H (*pmpH*) gene that occurs only in LGV strains (Morré *et al.*, 2005) (**Chapter 6**). The inclusion of RNase P in the assay is to detect any inhibition in the PCR and to ensure that each sample contains human DNA (as would be expected in the correct sample types). At NWLP, all residual DNA extracts are stored at -20°C.

Clinical specimens from Burrell Street Sexual Health Centre are tested for *C. trachomatis* at Viapath, a private-sector diagnostic laboratory based at St Thomas' Hospital. Viapath also provides *C. trachomatis* testing for clinical specimens collected at Streatham Hill Clinic, Walworth Road Clinic, and Harrison Wing. At Viapath, clinical specimens are tested for *C. trachomatis* using the Aptima *C. trachomatis/N. gonorrhoeae* (CT/NG) Combo 2 Assay (*Hologic®, USA*). This assay is a NAAT that detects *C. trachomatis* ribosomal RNA (rRNA) in clinical specimens (Hologic, 2017). At Viapath, all NAAT-positive specimens from MSM are subsequently tested for LGV. Nucleic acids are isolated from specimens using the Complex 200 Protocol for the QIAsymphony DSP Virus/Pathogen kit (QIAGEN, USA) and LGV confirmation testing is undertaken using an adapted version (adaptations to the primers and probes as indicated in Appendix 5) of the triplex LGV PCR assay developed at STBRL (**Appendix 5**). Residual DNA extracts are stored at -20°C.

Ethical approval was sought to collect residual *C. trachomatis* DNA extracts confirmed as LGV ("LGV-positive") from NWLP and Viapath. As NWLP and Viapath each receive clinical specimens from several GUM clinics across London, we sought ethical approval to collect DNA extracts from all clinics providing specimens to NWLP and Viapath. These were: 56 Dean Street, John Hunter Clinic, 10 Hammersmith Broadway, Jefferiss Wing, Burrell Street Sexual Health Centre, Streatham Hill Clinic, Walworth Road Clinic, and Harrison Wing. The ethical approvals process required to obtain these DNA extracts is outlined in section 5.4. The inclusion criterion for the study was: all residual

DNA extracts confirmed as LGV by *pmpH* PCR that were sourced from patients aged >16 years that attended the aforementioned clinics. We opted to exclude any LGV-positive DNA extracts collected from patients aged <16 years from the study, as inclusion of these specimens would have required parental permission. This would not logistically have been feasible given the scale of the genotyping study.

As the study was designed as a preliminary investigation with the purpose of assessing genotypic diversity in LGV strains; we elected not to collect patient clinical data, instead focusing solely on genotyping the DNA extracts and forming conclusions on the genotypic diversity within the LGV strains. Whilst not collecting clinical data is a limitation of the study and this is addressed in section 5.10; it does give some scope for a future multi-city genotyping survey of LGV in the UK, which could specifically analyse clinical data in addition to genotypic data (**5.10**).

Clinical specimens were tested at NWLP and Viapath, and all LGV-positive DNA extracts were set aside for the study and stored at -80°C. The window for sample collection was February 2018 – June 2019. This sample collection window allowed us to collect as many LGV-positive DNA extracts for the study as possible. During this time, a total of 230 LGV-positive DNA extracts were collected for the study; 180 from NWLP and 50 from Viapath. This represented approximately 12% (230/1,880) of all positive LGV tests in England during 2018 and 2019.

For the study, all DNA extracts confirmed as LGV at NWLP and Viapath were selected consecutively with no bias for positivity strength. DNA extracts were anonymised by laboratory personnel at NWLP and Viapath and labelled with “NWLP” or “V” followed by a sample number i.e. 001. Prior to anonymisation, laboratory personnel screened the DNA extracts to exclude any DNA extracts sourced from individuals aged <16 years. DNA extracts were transported on dry ice to the University of Southampton. Upon arrival, DNA extracts were stored at -20°C.

5.4 Ethical Approvals

Ethical approvals were required to gain access to the residual LGV-positive DNA extracts from NWLP and Viapath. The ethical approvals process is outlined in Figure 19. All research studies carried out at the University of Southampton are screened and approved by University of Southampton Ethics and Research Governance Online (ERGO) before external ethical approval is sought. Once ERGO has issued an outcome of approval, the University of Southampton provides insurance and sponsorship for the study. The ERGO reference number for this study was 45840. The study required access to DNA extracts collected from NHS GUM clinic patients, hence ethical approval was required from a research ethics committee (REC). REC approval was sought by submitting an Integrated Research Application Service (IRAS) online application. This study (IRAS reference number: 254542) was allocated for proportionate review by the East of Scotland REC. Proportionate review is an accelerated ethical review process for studies raising no material ethical concerns (i.e. minimal risk, burden or intrusion to participants). The study was approved by REC (reference number: 19/ES/0012), and subsequently by the Health Research Authority (HRA), with the condition that management approvals be sought from each NHS Trust involved in the study. Management approvals were subsequently obtained from Imperial NHS Foundation Trust (NWLP) and Guy's and St Thomas' NHS Trust (Viapath). Material Transfer Agreements (MTAs) were prepared by the University of Southampton legal department. These MTAs governed the transfer of the LGV-positive DNA extracts from the testing sites (Viapath and NWLP) to the University of Southampton.

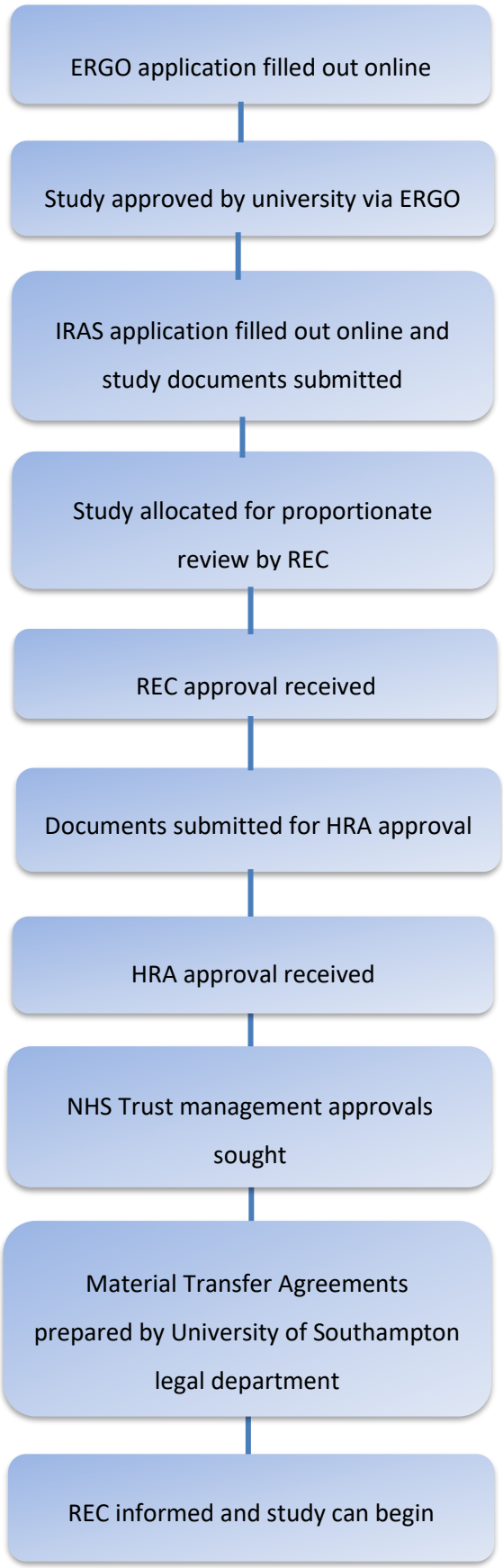


Figure 19 Flow diagram showing the process required to gain ethical approval for this study.

5.5 Clinical DNA Extracts

A total of 230 DNA extracts that were confirmed as LGV-positive using the *pmpH* real-time PCR-based assay at NWLP and Viapath were obtained for this study. These included 180 DNA extracts from NWLP and 50 DNA extracts from Viapath. The volumes of DNA extracts received were variable – the average volume of an NWLP extract was 53 µL, and the average volume of each extract from Viapath was 23 µL. Fifty per cent of the volume of each extract was aliquotted into sterile microtubes prior to commencing genotyping, to form a "reproducibility stock". The purpose of a "reproducibility stock" was to ensure that there was a set of un-opened DNA extracts in case of future analyses, for any confirmatory tests, or if required for publication. Sterile microtubes were labelled with the matching sample number and "RS" (for reproducibility stock), and the reproducibility stock was stored in a separate -20°C freezer to the original DNA extracts.

The DNA extracts included in the study were collected from patients between February 2018 and June 2019 (**Appendix 6**). All DNA extracts from NWLP were sourced from rectal swabs. The GUM clinic distribution of DNA extracts from NWLP was as follows: 56 Dean Street (n= 151, 84%), 10 Hammersmith Broadway (n= 11, 6%), Jefferiss Wing (n= 7, 4%), John Hunter Clinic (n= 6, 3%), and unknown (n= 1). The four remaining DNA extracts from NWLP were External Quality Assurance (EQA) samples (**Appendix 6**). EQA samples are quality control samples that are included in the diagnostics process to assure lab quality. They are typically panels of known control material sent out blind to labs to test their quality.

The 50 DNA extracts from Viapath were sourced from 23 pooled "3-in-1" specimens, 24 rectal swabs, 1 throat swab and 2 urines (**Appendix 6**). The GUM clinic distribution of DNA extracts from Viapath could not be obtained as this information was purged from the Lab Information System at Viapath.

5.6 Typing of LGV-positive DNA extracts from London

5.6.1 Methods used to assign MLVA-*ompA* genotypes to the DNA extracts

The DNA extracts were genotyped using the Pedersen *et al.* MLVA-*ompA* genotyping system (2008). The three VNTR regions (CT1335, CT1299, and CT1291) and the *ompA* gene were amplified from the DNA extracts using PCR (**2.9**). The PCRs were conducted in 20 µL volumes consisting of 10 µL Phusion Flash High-Fidelity DNA Polymerase, 0.5 µM of forward and reverse primers (**Table 3** and **5.6.1.2**), 1 µL neat DNA, and UltraPure™ DNase/RNase-Free distilled H_2O (**2.9**). The PCR conditions used are listed in Table 4 (**2.9**). The PCRs were conducted using appropriate controls (**5.6.1.1**). PCR products were subsequently loaded onto 2% agarose gels to check amplicon size and quality (**2.11**). Bands of the expected amplicon size (**Table 3**) were excised from gels, and DNA was purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega, UK)* (**2.10**). Purified PCR amplicons were commercially sequenced at Source Bioscience Cambridge (**2.13**). Alphabetical *ompA* genotypes were assigned using the method outlined in section 2.19.2. VNTR types were assigned by comparison with existing VNTR types (**Table 5**).

5.6.1.1 Controls

Positive and negative controls were employed in each PCR run. The positive control used was SotonK1 (Genbank Accession No. HE601794.1) at 1:100 dilution. SotonK1 is an *ompA* genotype K *C. trachomatis* strain that was isolated from the endocervix of a patient in 2009. The live isolate is stored in the Chlamydia Biobank (http://www.chlamydiabiobank.co.uk/Strainlists.html). Genomic DNA was previously isolated from SotonK1 using the NucleoSpin DNA Tissue Kit (**2.4.2**) and was stored at -20°C. SotonK1 was selected as the positive control so that it could be differentiated from the LGV-positive DNA extracts in the event of contamination. A 1:100 dilution of SotonK1 was used as this dilution produced bands of strong intensity after 40 cycles of amplification, and this dilution was previously used for MLVA-*ompA* genotyping by Labiran (2014).

No-template negative controls (NTC) were employed in each PCR run – the NTCs contained 1 µL of UltraPure™ DNase/RNase-Free distilled H_2O (**2.9**) in lieu of template DNA. Approximately 1 NTC per 10 DNA extracts was employed in each PCR run. This ratio of NTC: DNA extract was decided following standard protocols employed by diagnostic testing facilities. The use of multiple NTCs ensured that any contamination during the process could be detected quickly.

5.6.1.2 Primers

Amplification of VNTR regions in clinical DNA extracts was initially conducted using the primers outlined in Pedersen *et al.* (2008) and Wang *et al.* (2011).

In 2014, Labiran (2014) reported issues with the original primers used to amplify the VNTR regions (CT1335F, CT1335R, CT1299F, CT1299R, CT1291F, and CT1291R) (**Table 3**). In the Labiran study, VNTR types could be assigned to only 30% of extracts using these primers. Labiran designed a new set of primers to amplify the VNTR regions (CT1299F*, CT1299R*, CT1291F* and CT1291R*; **Table 3**). The new set of forward primers annealed upstream of the original VNTR amplicon sequences, and the reverse primers downstream, so that the original amplicon sequences were encompassed by the new set of primers. Labiran (2014) found that PCR amplification using the new primers increased the percentage of samples that could be genotyped to 76% (which is in line with the typeability of other MLVA-*ompA* studies) (Wang *et al.*, 2011); and visible amplicons were produced for samples that previously had not amplified using the original primer set. It was suggested that the DNA extracts that had previously failed to amplify the VNTRs in Labiran (2014) may have had single nucleotide polymorphisms (SNPs) in the primer binding sites that prevented the original primers from annealing.

Before embarking on the full study, we decided to select a small number of DNA extracts and to amplify one of the VNTR regions, and to see whether PCR amplicons could be produced using the original primer set. Five DNA extracts from NWLP (N001 – N005) were selected for PCR amplification of the CT1335 VNTR region using the original primers (CT1335F and CT1335R; **Table 3**). PCR reaction tubes containing Phusion Flash High-Fidelity Master-mix (**2.9**), CT1335F and CT1335R (**2.9.1, Table 3**), and 1 µL of neat DNA were prepared (**2.9**), and PCRs were conducted using the PCR conditions listed in section 2.9 (**Table 4**). The PCR products were subsequently loaded onto 2% agarose gels and imaged under UV light (**2.11.2**). One (N004) of the five DNA extracts produced a very faint band of the expected size (approx. 200 bp), and a band of the same size was present in the lane loaded with positive control. This indicated that the CT1335 region had been amplified in the extract and the positive control; however, the other four DNA extracts did not amplify at the CT1335 region. To avoid depleting the volume of the first five DNA extracts, it was decided to repeat the process with a different set of five DNA extracts (N006 – N010) in case the first five extracts were anomalies. When the agarose gel containing PCR products was visualised under UV light, a faint band of the expected size (~ 200 bp) was visible for one of the five DNA extracts (N006). The positive control produced a product of the same size (**Figure 20**).

Figure 15 A 2% agarose gel showing the PCR products after amplification of the CT1335 VNTR region in five LGV-positive DNA extracts from North West London Pathology (NWLP) (N006-N010), with appropriate controls. (+) represents the positive control, SotonK1, at 1:100 dilution; (-) represents the no template negative control (NTC). H2 represents the DNA ladder HyperLadder II.

It was decided to test the extended primers, CT1335F* and CT1335R*, to see if application of these primers could produce CT1335 PCR amplicons. We selected the first ten NWLP extracts (N001 - N010) and set up two PCR tubes containing Master-mix for each extract – one tube contained 1 µL of neat DNA, and the second tube contained 3 µL of neat DNA. This was to test whether a higher sample DNA concentration could result in PCR amplicons being produced if the reaction tubes containing 1 µL of neat DNA failed to amplify. Of the ten DNA extracts tested using CT1335F* and CT1335R*, nine DNA extracts produced amplicons of the expected size (for tubes containing 1 µL and 3 µL of DNA). The DNA concentration of the DNA extract that did not amplify at CT1335 (N007) was determined by fluorometry using a Qubit® (**2.12**). The concentration of DNA in N007 was 3.3 ng/µL. The DNA concentrations of N001 – N006 and N008 – N010 ranged from 15.8 ng/µL to 63.5 ng/µL. It is possible that the DNA extract that did not amplify contained too low an amount of *C. trachomatis* DNA to produce amplicons. It is important to note that this DNA concentration is the total DNA concentration of the sample, and not the *C. trachomatis* DNA concentration. As such, a higher total DNA concentration does not necessarily indicate a high *C. trachomatis* DNA concentration. It was decided to use the CT1335F* and CT1335R* primers to amplify all DNA extracts in the study.

This process was repeated for the other two VNTR markers, CT1299 and CT1291; however to conserve sample volumes, we used 1 µL of DNA. For CT1299, PCR reaction tubes containing the CT1299F* and CT1299R* primers, master-mix and 1 µL of DNA, were set up for five DNA extracts.

All five DNA extracts produced PCR amplicons using this set of primers. It was decided to use the CT1299F* and CT1299R* primers to amplify all DNA extracts in the study. For CT1291, reaction tubes were set up containing the original CT1291F and CT1291R primers, master-mix and 1 µL of DNA. PCR amplicons were produced for all but one DNA extract. As a result, the CT1291F and CT1291R primers were used for amplification of all DNA extracts in the study. The PCR conditions remained unchanged (**Table 4**).

For each DNA extract, after purification of PCR amplicons and sequencing; sequence chromatograms were visualised. Any DNA extracts that produced amplicons that yielded poor quality sequences were repeat-sequenced by Source Bioscience. If there was no improvement in sequence quality, these DNA extracts were repeat PCR amplified.

5.7 Genotyping Results

5.7.1 *OmpA* distribution

In total, we received 226 DNA extracts from patients. Of these, 170 (75.2%) were assigned an *ompA* genotype in this study (**Appendix 7**). The remaining 56 DNA extracts that were not assigned an *ompA* genotype in this study did not produce amplicons even upon repeat PCR. One hundred and sixty two extracts (95.3%) were assigned an LGV *ompA* genotype, and 8 (4.7%) were assigned non-LGV *ompA* genotypes. The non-LGV *ompA* genotypes identified were genotypes E (n= 3), G (n= 4), and J (n= 1). All LGV *ompA* genotypes were detected in the sample set, except L3. The most prevalent LGV *ompA* genotype identified in the study was L2/434/Bu (n= 78, 48.1%), followed by L2b/UCH-1/proctitis (n=48, 29.6%) (**Table 21**). We also detected three extracts with *ompA* genotype L1 (n= 3, 1.9%). We detected four LGV *ompA* variants. These included a hybrid *ompA* sequence, L2b/D (n= 8, 4.9%) (**Table 21**). This hybrid L2b/D *ompA* sequence is detailed in section 5.7.4. We also detected the L2 *ompA* variant L2h (Genbank Accession No. MH253042) (n= 1, 0.62%) (**Table 21**); in addition to the L2b *ompA* variants L2bV5 (Genbank Accession No. MH253040) in 18 extracts (11%) (**Table 21**); and L2bV6 (Genbank Accession No. MH253042) in 6 extracts (3.7%) (**Table 21**).

The distribution of *ompA* genotypes identified in the London study population is shown in Figure 21.

We also received four external quality assurance (EQA) samples from NWLP. These are indicated by asterisks in Appendix 6 and 7. The *ompA* genotypes assigned to these samples were: L2/434/Bu (n=2), G (n=1), and one sample could not be assigned an *ompA* genotype. The EQA sample that could not be assigned an *ompA* genotype did not produce amplicons even upon repeat PCR. This sample also did not amplify at any of the VNTR loci.

Table 21 LGV *ompA* sequences derived from patient* LGV-positive DNA extracts from London

LGV *ompA* sequence (Genbank Accession Number)	**n**	**% of 162 patient* DNA extracts assigned an LGV *ompA* genotype**
L2/434/Bu (AM884176)	78	48.1
L2b/UCH-1/proctitis (AM884177)	48	29.6
L2bV5 (MH253040) (most similar to L2b/UCH-1/proctitis, with C493A and G271A)	18	11.0
L2bV6 (MH253041) (most similar to L2b/UCH-1/proctitis, with G998A)	6	3.7
L2h (MH253042) (most similar to L2/434/Bu, with A997G)	1	0.62
L2b/D (MN094864.1)	8	4.9
L1/440/LN (HE601950)	3	1.9

**OmpA* genotypes from clinical specimens collected from patients were included in the Table. *OmpA* genotypes assigned to EQA samples (i.e. non-patient samples) are excluded from the Table, but these were assigned the following: L2/434/Bu (n=2), G (n=1), and un-typed (n=1).

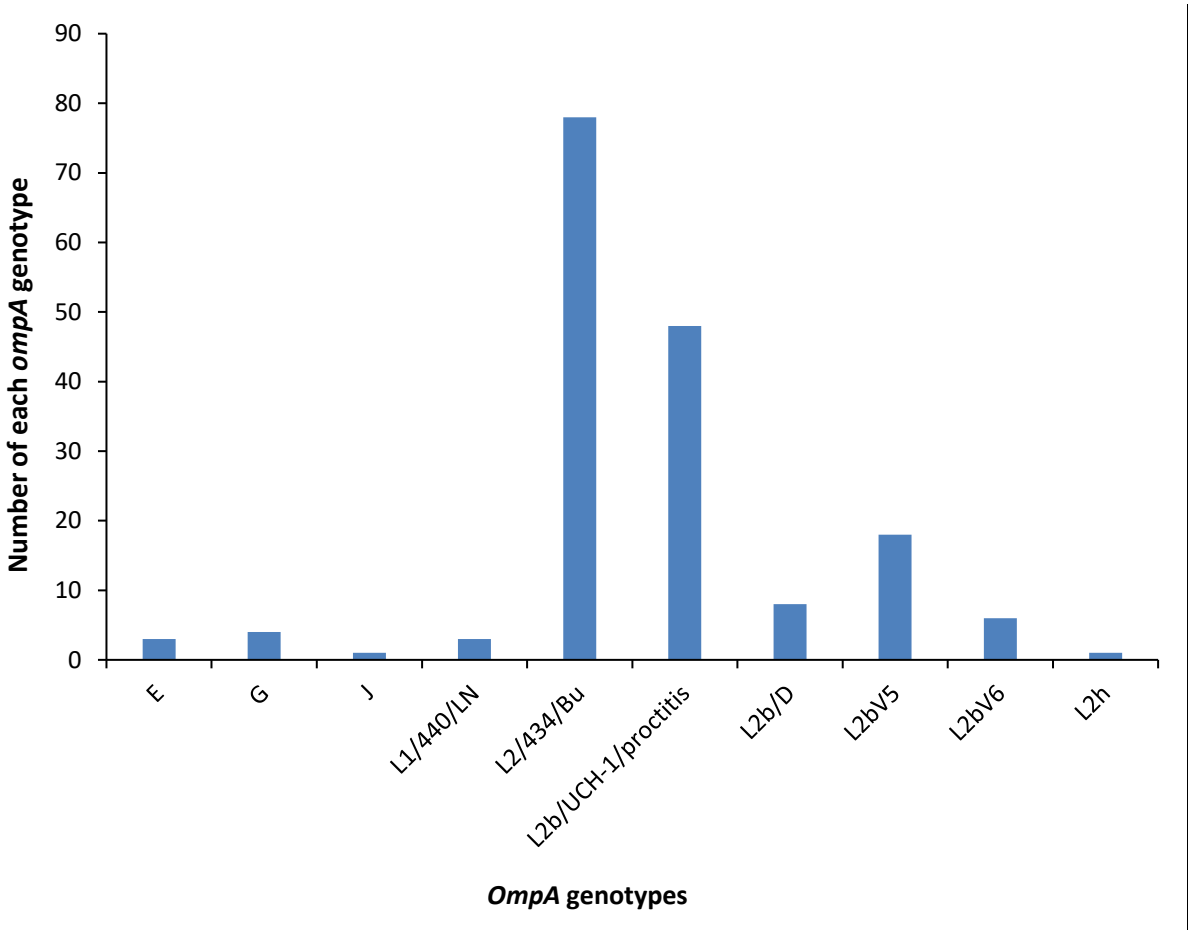


Figure 21 The distribution of *ompA* genotypes within LGV-positive DNA extracts from patients*in London. LGV *ompA* genotypes were assigned to 162 clinical DNA extracts, and non-LGV *ompA* genotypes were assigned to 8 patient DNA extracts. These included *ompA* genotypes E (n=3), G (n= 4), and J (n= 1).

*These are *ompA* genotypes that were derived from clinical specimens collected from patients. We also obtained four External Quality Assurance (EQA) samples that were assigned the *ompA* genotypes: L2/434/Bu (n=2), G (n=1), and un-typed (n=1).

5.7.1.1 *OmpA* distribution by testing provider

In this study, DNA extracts from patients were provided by NWLP (n= 176) and Viapath (n = 50). Of the 176 patient DNA extracts provided by NWLP, 144 (81.8%) extracts were assigned *ompA* genotypes in the study (**Appendix 7**). Twenty six (52%) DNA extracts from Viapath were assigned *ompA* genotypes in the study (**Appendix 7**). The distribution of *ompA* genotypes identified in DNA extracts tested at NWLP and Viapath is shown in Figure 22.

Nine distinct *ompA* genotypes were detected in extracts from NWLP – six LGV *ompA* genotypes and three non-LGV *ompA* genotypes (*ompA* genotypes E, G and J). The most prevalent *ompA* genotype assigned to clinical DNA extracts from NWLP was L2/434/Bu (n= 62), followed by L2b/UCH-1/proctitis (n= 44). We also identified 7 DNA extracts with the L2b/D hybrid *ompA* sequence. This represented 88% of all DNA extracts assigned L2b/D in the study. Within extracts tested at NWLP, we detected the LGV *ompA* variants L2bV5 (n= 15), and L2bV6 (n= 6), representing 83% and 100% of DNA extracts assigned these genotypes in the study, respectively. The LGV *ompA* variant L2h was not identified amongst DNA extracts tested at NWLP (**Figure 22**). We also obtained four External Quality Assurance (EQA) samples from NWLP – these were assigned the *ompA* genotypes L2/434/Bu (n=2), G (n=1), and one could not be assigned an *ompA* genotype.

Of the extracts tested at Viapath, six distinct *ompA* genotypes were detected. These included five LGV *ompA* genotypes and one non-LGV *ompA* genotype (*ompA* genotype E in one extract). The most prevalent *ompA* genotype assigned to DNA extracts tested at Viapath was L2/434/Bu (n= 16), followed by L2b/UCH-1/proctitis (n= 4). The *ompA* genotype L2/434/Bu represented 62% of all DNA extracts from Viapath assigned an *ompA* genotype. Within DNA extracts tested at Viapath, we detected one extract with the L2b/D hybrid *ompA* sequence. We also detected the LGV *ompA* variants L2bV5 in three extracts from Viapath, representing 17% of extracts with this variant in the study. No DNA extracts tested at Viapath with the L2bV6 variant were detected. One extract from Viapath had an identical *ompA* sequence to the LGV *ompA* variant L2h (Genbank Accession No. MH253042).

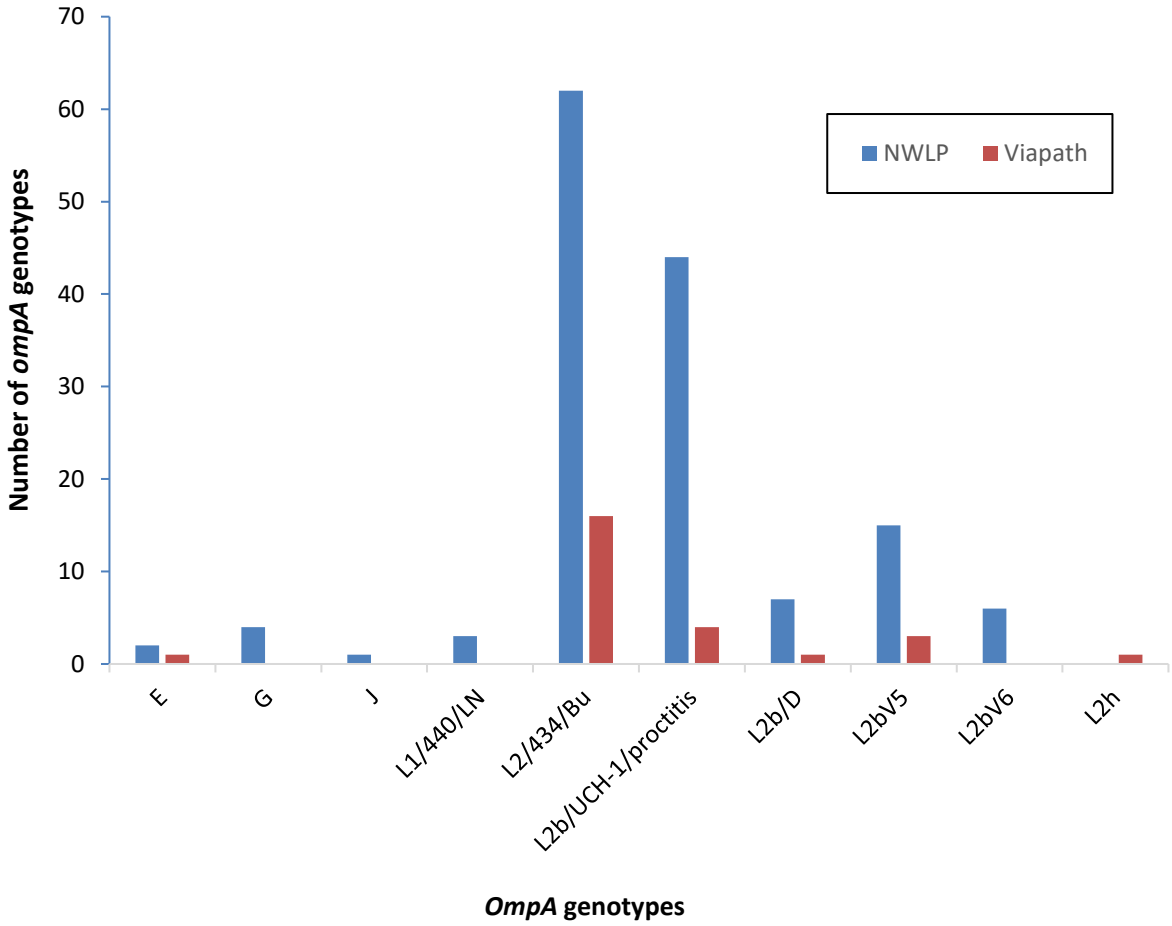


Figure 22 The distribution of *ompA* genotypes within LGV-positive clinical DNA extracts tested at North West London Pathology (NWLP) and Viapath. This does not include the four External Quality Assurance (EQA) samples that were obtained from NWLP – these were assigned *ompA* genotypes L2/434/Bu (n=2), G (n=1), and one could not be assigned an *ompA* genotype.

5.7.2 MLVA-*ompA* genotype distribution

Of the 226 patient DNA extracts genotyped in the study, 180 (79.6%) DNA had a type assigned at all three VNTR markers. One hundred and fifty nine (70.4%) DNA extracts were assigned complete MLVA-*ompA* genotypes in the study. Of the 67 remaining extracts with incomplete MLVA-*ompA* genotypes, 38 (41.2%) had a type assigned to at least two markers, 5 (7.5%) had a type assigned to one marker, and 24 (35.8%) were unable to be typed at any of the four MLVA-*ompA* markers (**Appendix 7**).

One hundred and fifty seven (98.7%) of the 159 extracts assigned complete MLVA-*ompA* genotypes were assigned LGV MLVA-*ompA* genotypes, and the remaining 2 extracts were assigned 3.9.3 - G.

Of those assigned LGV MLVA-*ompA* genotypes, ten distinct MLVA-*ompA* genotypes were identified (**Table 22**). The most prevalent LGV MLVA-*ompA* genotype assigned in the study was 1.9.2b - L2 (n= 75, 47.8%). The MLVA profiles 1.9.2b and 1.9.3b were each associated with several *ompA* genotypes: 1.9.2b was assigned to extracts with *ompA* genotypes L2, L2b, L2b/D, and the LGV *ompA* variants L2bV5 and L2h (**Table 22**); and 1.9.3b was assigned to extracts with *ompA* genotypes L2b, L2bV5 and L2bV6 (**Table 22**).

All eight DNA extracts with the hybrid L2b/D *ompA* sequence were assigned the MLVA-*ompA* genotype 1.9.2b-L2b/D. The DNA extract assigned the MLVA profile 1.4a.2b was *ompA* genotype L2b. All of the five DNA extracts containing the L2bV6 *ompA* variant were assigned the MLVA-*ompA* genotype 1.9.3b - L2bV6 (**Table 22**). DNA extracts containing the L2bV5 *ompA* variant were divided into two sub-types: 1.9.2b (56%) and 1.9.3b (44%). All *ompA* genotype L2 DNA extracts were assigned the MLVA profile 1.9.2b (**Table 22**).

Six extracts with partial MLVA-*ompA* genotypes and with a non-LGV *ompA* genotype were also identified in the study (**Appendix 7**). These included: 3.X.3-G (n= 2), 3.X.3- J (n= 1), 8.5.X-E (n=1), and 8.6.X-E (n= 2) (where X indicates the marker that could not be assigned a VNTR type) (**Appendix 7**).

We also received four External Quality Assurance (EQA) samples from NWLP. Three of the four samples (75%) were assigned a complete MLVA-*ompA* genotype in the study – 1.9.2b-L2 (n=2), and 3.9.3-G (n=1). One EQA sample could not be assigned a complete MLVA-*ompA* genotype. This sample could not be assigned a VNTR type at any of the three VNTR loci, nor could it be assigned an *ompA* genotype. No amplicons were produced at any of the VNTR loci or *ompA* for this sample, even upon repeat PCR and re-sequencing.

Table 22 Complete LGV MLVA-*ompA* genotypes identified in the study (n=157). Extracts with partial LGV MLVA-*ompA* profiles were excluded from the table. Also excluded from the table were the two patient DNA extracts assigned the non-LGV MLVA-*ompA* genotype, 3.9.3-G, and the 3 EQA samples that could be assigned complete MLVA-*ompA* genotypes (1.9.2b-L2 (n=2), and 3.9.3-G (n=1)).

***OmpA* (Genbank Accession No.)**	**MLVA†**	**n (% of 157 extracts)**
L1/440/LN (HE601950)	5.9.2b	3 (1.9)
L2/434/Bu (AM884176)	1.9.2b	75 (47.8)
L2b/UCH-1/proctitis (AM884177)	1.4a.2b 1.9.2b 1.9.3b	1 (0.64) 43 (27.4) 3 (1.9)
L2b/D (MN094864.1)	1.9.2b	8 (5.1)
L2bV5 (MH253040)	1.9.2b 1.9.3b	10 (6.4) 8 (5.1)
L2bV6 (MH253041)	1.9.3b	5 (3.2)
L2h (MH253042)	1.9.2b	1 (0.64)

† MLVA profile was designated by the 3 VNTR loci in the order: CT1335; CT1299; and CT1291.

5.7.3 VNTR sequence variants identified in the study

5.7.3.1 Sequence variants identified in extracts assigned LGV *ompA* genotypes

Table 23 shows the distribution of VNTR sequence variants identified in clinical DNA extracts assigned LGV *ompA* genotypes (n= 162). In total, 157 (96.9%) of 162 extracts assigned a CT1335 type were CT1335 type 1 (GAAAAAG-**9T8A**-GCTTTTGT) (**Table 23)**. The remaining three extracts assigned a CT1335 type were CT1335 type 5 (GAAAAAG-**11T8A**-GCTTTTGT) (**Table 23**), of which all were *ompA* genotype L1 and were assigned the complete MLVA-*ompA* genotype 5.9.2b -L1 (**Appendix 7**).

Also detected in the study were VNTR variants first identified in Wang *et al.* (Wang *et al.*, 2011): CT1299 type 4a (TTTTTATTCT-**10C**-T3C-ATCAAA) (n= 1), CT1299 type 9 (TTTTTATTCT-3C2T-**6C**-ATCAAA) (n= 159), and CT1291 type 2b (AAAATAGTCTA-**8C**-TATTG) (n= 144) (**Table 23**). The CT1299 type 9 was assigned to 159 (98.1%) of the 162 DNA extracts with LGV *ompA* genotypes (**Table 23**). This variant was detected in extracts with *ompA* genotypes L1 (n= 3), L2 (n= 78), L2b (n= 46), L2b/D

(n= 8), L2h (n= 1), L2bV5 (n= 18), and L2bV6 (n= 5). The extract assigned CT1299 type 4a in the study had the complete MLVA-*ompA* genotype 1.4a.2b-L2b (**Appendix 7**). All extracts assigned a CT1291 type 2b in the study were the 2b variant identified in Wang *et al.* (2011).

The VNTR variant code, CT1291 type 3b (AAAATAGTCTA-**9C**-TATTG), was identified in 20 extracts in the study (**Appendix 7**). Sixteen (80%) extracts with the 3b variant could be assigned an *ompA* genotype, of which five extracts were L2bV6 (31.3%), eight extracts were L2bV5 (50%), and three extracts were L2b (18.8%) (**Table 23**). The MLVA profile of these sixteen extracts was 1.9.3b. The CT1291 type 3b was previously described by Satoh *et al.* (2014), where it was assigned to the LGV reference strains L1/440/LN and L2/434/Bu.

Table 23 VNTR sequence analysis of patient DNA extracts assigned LGV *ompA* genotypes (n=162). VNTR type codes of extracts that were assigned non-LGV *ompA* genotypes (n=8) or that could not be assigned an *ompA* genotype (n=56), were excluded from the table.

VNTR type codes	**n**	**CT1335 Variants**[a]	**References**
1	157	GAAAAAG-**9T8A**-GCTTTTGT	(Pedersen *et al.*, 2008)
5	3	GAAAAAG-**11T8A**-GCTTTTGT	(Pedersen *et al*., 2008)
Non-typeable	2		
		CT1299 Variants[a]	
4a	1	TTTTTATTCT-**10C**-T3C-ATCAAA	(Wang *et al.*, 2011)
9	159	TTTTTATTCT-3C2T-**6C**-ATCAAA	(Wang *et al.*, 2011)
Non-typeable	2		
		CT1291 Variants[a]	
2b	144	AAAATAGTCTA-**8C**-TATTG	(Wang *et al.*, 2011)
3b	16	AAAATAGTCTA-**9C**-TATTG	(Satoh *et al.*, 2014)
Non-typeable	2		

[a] Repeating mononucleotide sequences at each VNTR locus are shown in bold

Flanking region variation is underlined.

5.7.4 Detection of a hybrid L2b/D *ompA* genotype

In this study, eight *ompA* sequences with an L2b/D hybrid profile were identified (**Figure 21**). In each of these eight extracts, the first 365 bp (numbers given relate to L2b/UCH-1/proctitis) of *ompA* were identical to *ompA* L2 and L2b reference sequences (L2/434/Bu and L2b/UCH-1), whilst the region spanning 366 – 1,023bp revealed an *ompA* genotype D profile matching the reference strain D/UW-3/CX (Genbank Accession No. NC_000117.1) (**Appendix 8**). A nucleotide BLAST of the eight sequences indicated 100% sequence identity to the *ompA* sequence of a hybrid L2b/D strain identified in Portugal (Genbank Accession No. MN094864.1) (Borges *et al.*, 2019). An alignment of the *ompA* sequence of the eight extracts with the Portuguese L2b/D *ompA*, and *ompA* of L2/434/Bu, L2b/UCH-1/proctitis, and D/UW-3/CX is shown in Appendix 8. In this study, the eight hybrid sequences were designated L2b/D, in order to distinguish them from DNA extracts with *ompA* sequences matching L2b/UCH-1/proctitis (which were assigned *ompA* genotype L2b).

All eight DNA extracts assigned *ompA* genotype L2b/D could be assigned a complete MLVA-*ompA* genotype – of which all were 1.9.2b - L2b/D (**Table 21**). The VNTR profile of the Portuguese hybrid L2b/D strain (strain Ct_L2b/D_PT05; European Nucleotide Archive (ENA) Accession No. CAAKND010000000) was determined *in silico.* This was done by firstly converting the available .fastq files for the strain to .fasta using the conversion tool on the public server at usegalaxy.org (Afgan et al., 2018). Secondly, the .fasta files were opened using BioEdit (version 7.0.5.3), and the VNTRs located by inputting the VNTR primer sequences into the search tool. The VNTR profile of the Portuguese strain was: GAAAAAG-**9T8A**-GCTTTTGT at CT1335 (CT1335 type 1), TTTTTATTCT-3C2T-**6C**-ATCAAA at CT1299 (CT1299 type 9), and AAAATAGTCTA-**8C**-TATTG at CT1291 (CT1291 type 2b). This corresponded to a MLVA-*ompA* genotype of 1.9.2b-L2b/D – the same MLVA-*ompA* genotype as identified for the L2b/D extracts in the study.

5.8 Typeability of the MLVA-*ompA* genotyping system

The typeability of a genotyping system (**2.20.2**) is its ability to assign a genotype to all isolates tested by the genotyping system. The typeability of each MLVA-*ompA* genotyping marker, and the overall typeability of the MLVA-*ompA* system was calculated using the method detailed in section 2.20.2. The typeability of each of the markers for the patient clinical DNA extracts was as follows: CT1335 (86.7%, 196/226), CT1299 (85.4%, 193/226), CT1291 (83.6%, 189/226), and *ompA* (75.2%, 170/226). The overall typeability of the MLVA-*ompA* genotyping system in this study was 70.4% (159/226). Inclusion of the four External Quality Assurance (EQA) samples obtained from NWLP results in the following typeabilities: CT1335 (86.5% , 199/230), CT1299 (85.2% , 196/230), CT1291 (83.5% , 192/230), and *ompA* (75.2% , 173/230).

5.9 Simpson's Index of Diversity for the MLVA-*ompA* genotyping system

The Simpson's Index of Diversity was calculated using the Hunter-Gaston modification (Hunter and Gaston, 1988) to determine the discriminatory power of MLVA-*ompA* for the 226 DNA extracts that comprised the LGV genotyping study. We wished to determine the level of discrimination provided by the MLVA-*ompA* genotyping system, and to determine whether the addition of the three VNTR markers increased the discriminatory power of the system, compared to typing the extracts using just *ompA*. To do this, the equation provided in section 2.20.1 was used to calculate the discriminatory power of each of the four markers. This was then followed by calculating the discriminatory power of the three VNTR markers combined. Finally, the discriminatory power of the three VNTR loci in conjunction with *ompA* was calculated. For the 226 clinical DNA extracts, the following values were obtained: the marker CT1335 had a Simpson's Index of Diversity of 0.33; the marker CT1299 had an Index of Diversity of 0.29; the marker CT1291 had an Index of Diversity of 0.45; and *ompA* had an Index of Diversity of 0.77. The three VNTR markers had a combined Index of Diversity of 0.50. The MLVA-*ompA* genotyping system had a Simpson's Index of Diversity of 0.76 for this sample set. Inclusion of the four External Quality Assurance (EQA) samples obtained from NWLP resulted in the following values for discriminatory power for each of the four markers: the marker CT1335 had a Simpson's Index of Diversity of 0.33 (this remains unchanged); the marker CT1299 had an Index of Diversity of 0.29 (this remains unchanged); the marker CT1291 had an Index of Diversity of 0.45 (this remains unchanged); and *ompA* had an Index of Diversity of 0.77 (this remains unchanged). The three VNTR markers had a combined Index of Diversity of 0.50 (this remains unchanged). The MLVA-*ompA* genotyping system had a Simpson's Index of Diversity of 0.76 (this remains unchanged).

To more directly compare the findings of this study with that of Labiran *et al.* (2016), Simpson's Index of Diversity was calculated for each locus (CT1335, CT1299, CT1291, and *ompA*), in addition to the three VNTR loci combined, and for the MLVA-*ompA* genotyping system overall. To ensure that the results were more comparable, only those specimens with LGV *ompA* genotypes were included in the calculations. All non-LGV specimens were excluded. Table 24 indicates the values obtained in each study for Simpson's Index of Diversity.

Table 24 Simpson's Index of Diversity (D) values obtained for each locus (CT1335, CT1299, CT1291, and *ompA*), the three VNTR loci combined, and for the MLVA-*ompA* genotyping system, compared to those obtained by Labiran *et al.* (2016).

	D (CT1335)	D (CT1299)	D (CT1291)	D (*ompA*)	D (VNTRs)	D (MLVA-*ompA*)
This study (n=157)	0.04	0.02	0.15	0.33	0.33	0.69
Labiran (2016) (n=11)	0.56	0.55	0.78	0.32	0.93	0.95

5.10 Discussion

In 2016, Labiran *et al.* (2016) applied the MLVA-*ompA* genotyping system to genotype rectal swab specimens collected from men who have sex with men (MSM) in Brighton. The purpose of the Labiran study was to determine the distribution of MLVA-*ompA* genotypes in MSM with rectal *C. trachomatis* infections. The study identified eleven LGV specimens that were assigned *ompA* genotypes L2 (n = 1) and L2b (n = 9) (Labiran *et al.*, 2016). These specimens were resolved into nine distinct MLVA-*ompA* genotypes, with seven MLVA-*ompA* genotypes identified within the L2b specimens. These results demonstrated that the MLVA-*ompA* genotyping system was able to discriminate between LGV clinical specimens. In addition, the findings showed that a high level of genotypic diversity existed within the LGV strains circulating in Brighton. Whilst the study included only a small number of LGV clinical specimens, it hinted at the diversity that may exist within LGV strains in the UK. Prior to this Chapter, no large-scale high-resolution genotyping survey of LGV strains had been conducted in the United Kingdom. As a result, little was known about the nature of LGV strains circulating amongst MSM in the UK. This presents an issue given that the UK has the largest LGV epidemic in MSM in Europe (European Centre for Disease Control, 2019).

High-resolution genotyping systems can be used to determine the distribution of genotypes within populations (van Belkum *et al.*, 2007). This facilitates the prevalent strains circulating in a population to be identified, which can indicate the number of individuals in a population that are infected with each strain. In the UK, London accounts for 77% of all LGV cases in England (Public Health England, 2020d). It was decided to conduct a genotyping survey of LGV strains in London. The main aims of the study were to determine the distribution of LGV genotypes within MSM in London, in order to assess the level of genotypic diversity in this population. This would act as a pilot study, giving an idea of the diversity within LGV strains in the UK. The pilot study could be developed further in the future into a multi-city genotyping survey of LGV strains. It was demonstrated in Chapter 4 that the MLVA-*ompA* genotyping system had the highest typeability of

the available high-resolution genotyping systems when applied to 33 LGV strains of *C. trachomatis*. Chapter 4 also identified that MLVA-*ompA* was able to resolve the 33 LGV strains into eight distinct MLVA-*ompA* genotypes. It was decided to apply the MLVA-*ompA* genotyping system to genotype LGV-positive clinical specimens from London. An advantage of using the MLVA-*ompA* genotyping system to genotype LGV-positive specimens was that it enabled us to compare our genotyping study results with those collected in Brighton (Labiran *et al.*, 2016).

It was important to include as many LGV-positive clinical specimens in the study as possible, to ensure that the study was representative of a wide range of sexual networks in London. The LGV-positive DNA extracts genotyped in this study were obtained from two pathology providers, NWLP and Viapath. NWLP and Viapath provide *C. trachomatis* testing for clinical specimens collected from patients attending eight London-based GUM clinics. We obtained ethical approval to collect residual LGV DNA extracts stored at NWLP and Viapath. This was a strength of the study as it meant that our study included DNA extracts from patients from eight GUM clinics, hence is more likely to be representative of LGV strain diversity within the London population. The sole exclusion criterion for the study was that clinical specimens collected from patients younger than 16 years old were not to be included. The reason for this was because inclusion of specimens from patients under the age of 16 would have required permission from a parent or guardian, which was not feasible for this study. Given that very few cases of LGV are reported in this age-group, it is unlikely that many LGV specimens would have been collected from clinics (Public Health England, 2020d).

The LGV-positive DNA extracts in the study were collected from patients over an 18 month period (February 2018 - June 2019). In total, 230 LGV-positive DNA extracts were included in the study. However, it was identified that four of the DNA extracts from NWLP were external quality assurance (EQA) samples. EQA samples are quality control samples that are included in the diagnostics process to assure lab quality. They are typically panels of known control material sent out blind to labs to test their quality. The MLVA-*ompA* genotypes of the EQA samples were excluded from the LGV genotyping study. We obtained the following MLVA-*ompA* genotyping results from these four EQA samples: 1.9.2b-L2/434/Bu (n=2), 3.9.3-G (n=1), and one EQA sample could not be assigned a MLVA-*ompA* genotype. The sample that could not be assigned a MLVA-*ompA* genotype did not produce amplicons at any of the three VNTR loci, nor the *ompA* locus, even upon repeat PCR and sequencing. It is unclear as to why this sample did not produce amplicons during the study.

In accordance with a previous study (Cole *et al.*, 2020b), we identified that the L2b *ompA* genotype was not the most common in the pooled dataset; instead, the L2 *ompA* genotype predominated. The L2 *ompA* genotype was assigned to 48% of all LGV-positive DNA extracts genotyped in the study. Whilst the predominance of the L2 *ompA* genotype has been reported in Sweden (Isaksson

et al., 2017), Spain (Parra-Sánchez *et al.*, 2016), France (Peuchant *et al.*, 2016), and Austria and Croatia (Cole *et al.*, 2020b); this is the first report of its predominance within the United Kingdom. In a 2020 study that applied *ompA* genotyping to 22 LGV-positive rectal specimens from MSM in the UK; it was found that the L2b *ompA* genotype dominated – with 41.7% of extracts from the UK assigned *ompA* genotype L2b in the study, compared to one extract assigned *ompA* genotype L2 (8.3%). However, as noted, the Cole *et al.* study (2020b) included only a small number of LGV specimens, and as a result, the study results may not have been reflective of the entire population of LGV-infected MSM. In contrast, a retrospective study by Peuchant *et al.* (2016) applied *ompA* genotyping to 179 LGV-positive anorectal specimens collected from MSM in France between 2010 and 2015; identifying that L2b was the dominant genotype until 2012, when the proportion of cases caused by L2 increased.

We observed a high level of sequence diversity within the *ompA* gene, identifying a hybrid *ompA* genotype (L2b/D), two L2b *ompA* variants (L2bv5 and L2bv6), in addition to an L2 *ompA* variant (L2h) in the study. The L2bv5, L2bv6 and L2h LGV *ompA* variants were first identified in LGV-positive rectal specimens from the UK by Cole *et al.* (2020b), where they comprised 33% (n= 4), 4.5% (n= 1), and 4.5% (n= 1) respectively of the specimens assigned *ompA* genotypes. In our London study, the L2bV5 *ompA* variant comprised 11% (n = 18) of all DNA extracts assigned LGV *ompA* genotypes in the study. This variant has an *ompA* sequence that differs from L2b/UCH-1/proctitis by two SNPs (C493A and G271A). A second L2b *ompA* variant, L2bV6, was identified in six DNA extracts in the study. This variant has an *ompA* sequence that differs from L2b/UCH-1/proctitis by one SNP (G998A). Also identified in the study was one extract assigned *ompA* genotype L2h. The L2h *ompA* variant has an *ompA* sequence that is most similar to L2/434/Bu with one SNP (A997G). These SNPs are within variable segment 2 of the *ompA* gene, which is the primary antigenic marker of *C. trachomatis*. These results demonstrate that several LGV *ompA* variants are circulating in the London study population.

We identified three DNA extracts with *ompA* genotype L1 in the study. *OmpA* genotype L1 is not typically associated with the LGV proctitis epidemic in MSM, and is instead associated with the inguinal LGV syndrome (**1.6.3**). As this study did not collect any patient data relating to symptoms, we are unable to comment on the clinical presentation of the three patients with the L1 *ompA* sequence in the London study population.

The non-LGV *ompA* genotypes E (n= 3), G (n= 4) and J (n =1) were detected in this study. *OmpA* genotypes G and J have commonly been found in the rectum of MSM (Jeffrey *et al.*, 2010; Quint *et al.*, 2011), and co-infections with LGV and urogenital genotype E have been reported previously (Rodriguez-Dominguez *et al.*, 2015). However, the DNA extracts that were genotyped in this study

were those that had provided a positive test result in the LGV *pmpH* PCR assay at NWLP and Viapath, i.e. only extracts with the characteristic 36 base pair deletion within the *pmpH* gene. These non-LGV *ompA* genotypes had a low prevalence in the study, comprising 4.7% (8/170) of all DNA extracts assigned an *ompA* genotype. There are a few possible explanations for their detection in the study. Firstly, the extracts could have been false positives of the LGV biovar assay. The likelihood of this is slight, given that the assay has previously demonstrated excellent diagnostic performance in distinguishing LGV and non-LGV *C. trachomatis* infections in earlier studies (Chen *et al.*, 2007). A second possible explanation is that the individuals that provided the specimens assigned non-LGV *ompA* genotypes may have been infected with both an LGV strain of *C. trachomatis*, and a non-LGV strain. A study (Quint *et al.*, 2011) provided an estimate of the prevalence of non-LGV co-infections amongst MSM with LGV, and this was found to be 6.1%. In the Quint *et al.* study, of the six LGV cases with non-LGV strain co-infections, *ompA* genotypes E, G and J comprised 67%. Another possible explanation for the detection of the non-LGV *ompA* genotypes in our study, is that co-infections could have resulted in *pmpH* variants caused by genetic exchange between LGV and *ompA* genotype E, G, or J *C. trachomatis* strains. Recombination is a driver of evolution in *C. trachomatis* (Gomes *et al.*, 2007), and mixed infections can facilitate the selection of new recombinants, including the hybrid Portuguese strain (Borges *et al.*, 2019), which was caused by a recombination event between an L2b strain and a D strain; and the L2c strain identified by Somboonna *et al.* (2011), that was caused by a recombination between an L2 strain and a D strain.

Of the 226 DNA extracts obtained from patients in the study, 159 (70.4%) could be assigned complete MLVA-*ompA* genotypes. The most prevalent MLVA-*ompA* genotype assigned in the study was 1.9.2b - L2, which comprised 47.8% of all DNA extracts assigned complete MLVA-*ompA* genotypes in the study. This MLVA profile was associated with several *ompA* genotypes, including L2, L2b, L2b/D, and the LGV *ompA* variants L2bV5 and L2h. The MLVA profile 1.9.3b was also associated with more than one *ompA* genotype, and was detected in DNA extracts assigned *ompA* genotypes L2b, L2bV5, and L2bV6. This would suggest that several LGV *ompA* genotypes with the same VNTR profiles are co-circulating amongst MSM in London. This is likely the result of recombination within the *ompA* gene, as *ompA* is a recombination hotspot within the *C. trachomatis* genome (Gomes *et al.*, 2007). Our study showed that 100% of DNA extracts assigned *ompA* genotype L2 (for which a complete MLVA-*ompA* profile was obtained) had the MLVA-*ompA* genotype 1.9.2b-L2. This is indicative of a clonal spread of the 1.9.2b-L2 genotype within the London population (Harris *et al.*, 2012).

We noted the most diversity in MLVA-*ompA* genotypes within extracts designated *ompA* genotype L2b, with the VNTR profiles 1.9.2b, 1.9.3b, and 1.4a.2b assigned. Diversity in L2b genotypes was previously reported in Brighton by Labiran *et al.* (2016). However, whilst in London 1.9.2b - L2b and

1.9.3b - L2b comprised 43/157 (27%) and 3/157 (1.9%) respectively of extracts assigned a complete LGV MLVA-*ompA* genotype in the study; the Brighton study reported only three cases with 1.9.2b-L2b and no cases with 1.9.3b-L2b. In addition, the Brighton study identified eight LGV MLVA-*ompA* genotypes that were not detected in our London study population. These regional differences in MLVA-*ompA* genotypes in Brighton and London are likely the result of distinct dissemination patterns within each population, however, the differences could also represent a temporal shift in genotypes, given that our London study genotyped specimens collected in 2018 and 2019, and the Brighton study took place between 2011 and 2013. It should be noted that the Brighton study included only 11 LGV samples, that were assigned the *ompA* genotypes L2 (n=2) and L2b (n=9). Due to the small number of LGV samples included in the study, it could be suggested that the results of the Brighton study are not representative of the full picture of LGV strain diversity, hence caution should be applied when forming direct comparisons. This is discussed further on page 163.Our study identified the CT1291 type 3b variant (AAAATAGTCTA-**9C**-TATTG) in 20 DNA extracts. Sixteen of these DNA extracts were assigned a complete MLVA-*ompA* genotype, of which all were 1.9.3b-L2b. Prior to the London study, the CT1291 type 3b variant was detected by Satoh *et al.* (2014) within the L1/440/LN and L2/434/Bu reference strains, and these were assigned the MLVA-*ompA* genotypes 5.9.3b - L1 and 5.9.3b - L2. However, the Satoh *et al.* study did not include any L2b isolates, and the 3b variant was not detected in any of the 44 clinical isolates that were also MLVA-*ompA* genotyped by Satoh *et al.* As far as we are aware, our London study is the first report the CT1291 type 3b variant in L2b strains. We also detected VNTR variants that were first identified in Wang *et al.* (2011) and Pedersen *et al*. (2008). These included the CT1299 type 4a variant (TTTTTATTCT-**10C**-T3C-ATCAAA), CT1291 type 2b (AAAATAGTCTA-**8C**-TATTG), and CT1299 type 9 (TTTTTATTCT-3C2T-**6C**-ATCAAA). In our study, CT1299 type 9 comprised 159/162 (98.1%) of DNA extracts assigned an LGV *ompA* genotype. This contrasts with the findings of Wang *et al.* (2011), where CT1299 type 9 comprised only 6/93 (6.5%) of isolates yielding a complete MLVA-*ompA* genotype. It should be noted that the Wang *et al.* (2011) study genotyped only endocervical swabs collected from women and had a low sample size. The prevalence of the CT1299 type 9 variant in our study of LGV-positive DNA extracts could indicate that this variant has a closer linkage to LGV strains than genital *C. trachomatis* strains.

We detected a hybrid L2b/D *ompA* sequence in DNA extracts from NWLP (n= 7) and Viapath (n= 1). The first 365bp of the hybrid *ompA* sequence were identical to L2/434/Bu and L2b/UCH-1/proctitis reference sequences, whilst the region spanning 366bp – 1,023bp was identical to the *ompA* genotype D reference strain, D/UW-3/CX. The hybrid *ompA* sequence was found to be identical to the *ompA* of a recombinant L2b/D *C. trachomatis* strain detected in Portugal (Borges *et al.*, 2019). As of 2019, 25 cases with the recombinant L2b/D strain have been reported in Portugal. Our London

study is the first to report of the hybrid L2b/D *ompA* sequence in the United Kingdom. All eight extracts identified in the study with the L2b/D hybrid sequence were assigned the MLVA-*ompA* genotype 1.9.2b - L2b/D. We demonstrated by *in silico* means that the Portuguese hybrid L2b/D strain possessed the MLVA-*ompA* genotype 1.9.2b - L2b/D, confirming that the Portuguese L2b/D strain and the eight L2b/D extracts from the London study population shared the same MLVA-*ompA* genotype. Borges *et al.* (2019) performed whole genome sequencing of the Portuguese recombinant strain, and this revealed that the strain was the result of genetic transfer of a 4.2 kb fragment from a *C. trachomatis ompA* genotype D strain to an L2b strain. This recombinant fragment included 75% of the *ompA* gene encoding the major outer membrane protein (MOMP), in addition to four genes downstream of *ompA*, each with functional roles in protein synthesis (Borges *et al.*, 2019). The L2b/D *ompA* sequence has 19 amino acid replacements and one amino acid insertion compared to L2b/UCH-1/proctitis; and 18 of these amino acid replacements are within the surface-exposed epitope-enriched variable domains of the MOMP. This resulted in a new combination of epitopes, with mixed L2/L2b epitopes in variable domain 1 of the MOMP, and D epitopes in variable domains 2 and 4 of the MOMP (Borges *et al.*, 2019). The widespread genetic recombination within the hybrid L2b/D strain, particularly within the MOMP epitope region that is targeted by the host immune response, may have implications for the transmission and pathogenic capabilities of the strain. A particular limitation of our London study was that we did not collect any clinical data relating to patient symptoms, hence we cannot comment on the clinical manifestation of patients with the L2b/D *ompA* sequence in London. However, Borges *et al.* (2019) identified that all individuals with the hybrid strain presented with clinical features that are consistent with a typical LGV infection, including rectal pain, anal discharge, and rectal bleeding. It was noted that many of the individuals with the L2b/D strain in Portugal were involved in international sexual networks, which would explain how the variant likely reached the London population. Given the high-risk sexual practices of MSM (**1.8.3**), and the involvement of MSM in international sexual networks, it is very likely that this hybrid strain will be transmitted further within the high-risk population.

The MLVA-*ompA* genotyping method had limited resolution when applied to our London study population, with two LGV MLVA-*ompA* genotypes (1.9.2b - L2 and 1.9.2b - L2b) comprising 74% of DNA extracts assigned complete MLVA-*ompA* genotypes in the study. As a result, we did not reach a Simpson's Index of Diversity of 0.95, which is the Index value for a genotyping system to be considered to have more or less "ideal" resolution (van Belkum *et al.*, 2007). In fact, the three VNTRs combined had a Simpson's Index of Diversity of 0.50, which was lower than the level of discrimination provided by the *ompA* gene for the sample set (D = 0.77). The overall discriminatory index of the MLVA-*ompA* system was 0.76. This demonstrated that for this sample set, the highest

level of discrimination was provided by the *ompA* gene (0.77) followed by the combination of the *ompA* gene and the three VNTRs. However, even the VNTR with the highest discriminatory index in this study population (CT1291, D = 0.45) had a lower discriminatory capacity than the *ompA* gene for this sample set. To more accurately compare the findings of the London study with the Brighton study (Labiran *et al.,* 2016), Simpson's Index of Diversity was re-calculated for only those samples assigned an LGV *ompA* genotype in the London study (n=157) and in the Brighton study (n=11). This was because the values stated for Simpson's Index of Diversity in the Brighton paper were calculated for specimens assigned LGV *ompA* genotypes in addition to non-LGV *ompA* genotypes, which does not reflect the genotypic diversity of LGV strains only. We showed that Simpson's Index of Diversity for the Brighton study was higher at all four loci, in addition to the combined VNTR loci and the MLVA-*ompA* genotyping system overall, when compared to the London study. However it should be noted that the Labiran study consisted of only 11 LGV specimens, hence caution should be applied when directly comparing the results of both studies. As such, the lower discriminatory capacity of the MLVA-*ompA* system in our substantially larger London study is likely reflective of the clonal nature of the LGV epidemic in MSM.

The study was able to assign complete MLVA-*ompA* genotypes to 159 (70.4%) of DNA extracts included in the study. This was a slightly lower typeability than the Labiran *et al.* study in Brighton, which was able to assign MLVA-*ompA* genotypes to 76% of specimens. However, the lower typeability of MLVA-*ompA* in the London study population is not necessarily a limitation of the typing system applied to genotype the DNA extracts. Whilst typeability is a good indicator of the ability of a typing system to genotype clinical specimens; DNA extracts that failed to amplify and could not be assigned genotypes in the London study population may have contained degraded DNA or had a low *C. trachomatis* DNA concentration; hence would not have generated genotypes regardless of the genotyping system applied.

5.11 Conclusions

This was the first large-scale genotyping survey of LGV strains using the MLVA-*ompA* genotyping system that has ever been conducted in the United Kingdom. The study has demonstrated that the predominant *ompA* genotype within the London population is the L2 *ompA* genotype, and not the L2b *ompA* genotype that has been reported in UK populations since 2005 (Macdonald *et al.*, 2005). We have also provided evidence for the co-circulation of two dominant VNTR types in London – 1.9.2b and 1.9.3b – that are associated with several distinct *ompA* genotypes. We have reported a high level of diversity within *ompA* in the London study population, detecting several L2b and L2 *ompA* variants in London, including the L2bV5 and L2bV6 *ompA* variants previously identified in four samples from the UK in a 2020 study. We also provided the first UK report of a hybrid L2b/D

ompA profile previously detected in Portugal. At present, little is known about the pathogenic capability of this hybrid strain, and as a result, the public health implications of the L2b/D recombinant strain are as of yet unknown. We also demonstrated that the MLVA-*ompA* genotyping system had low resolution and the discriminatory capacity of the system for this sample set was low.

A limitation of our study was that we did not collect patient clinical data relating to symptoms, STI history, and HIV-positivity status. As a result, we were unable to form conclusions relating to the distribution of MLVA-*ompA* genotypes circulating within these high-risk sub-populations in London.

CHAPTER 6 Genomic Diversity within LGV *C. trachomatis* Strains Isolated from Asymptomatic and Symptomatic MSM in Brighton

Note: Bri038, Bri088, and Bri098 were grown in cell culture by Dr Colette O'Neill.

6.1 Introduction

As discussed in Chapter 5, *C. trachomatis* strains with *ompA* genotypes L1 – L3 cause Lymphogranuloma venereum (LGV). These strains are more invasive than those with *ompA* genotypes A – K; causing systemic infections, infecting monocytes and macrophages, before disseminating to local lymph nodes (**1.6.3**).

In 2003, an outbreak of LGV in MSM in the Netherlands was reported (Götz *et al.*, 2004). The majority of these patients presented with painful haemorrhagic proctitis, rectal discharge and rectal pain. These symptoms differed from the classical LGV inguinal syndrome seen in tropical regions, characterised by inguinal buboes and genital ulcers (**1.6.3**). It was speculated that the LGV proctitis outbreak represented a new emerging infection, and this appeared to be supported by sequence data of the polymorphic *ompA* gene, which showed that the LGV proctitis cases were caused by a new L2 *ompA* variant termed L2b (Spaargaren, Fennema, *et al.*, 2005). The L2b *ompA* variant is distinguished from L2 by a single SNP (A485G) that causes an N162S amino acid substitution within variable domain 2 of the major outer membrane protein (MOMP) (Seth-Smith *et al.*, 2021) (**1.11**). It was later shown that isolates collected from MSM in San Francisco between 1979 and 1985 contained the L2b *ompA* variant (Spaargaren, Schachter, *et al.*, 2005). This demonstrated that the newly observed proctitis cases were the result of a slowly-evolving LGV epidemic, rather than a new epidemic.

Since 2003, cases of LGV caused by the L2b variant have been reported amongst MSM globally. It is now known that the L2b epidemic is the result of clonal expansion and widespread transmission of the L2b variant amongst MSM (Harris *et al.*, 2012). The L2b strain is associated with both symptomatic and asymptomatic disease. Asymptomatic LGV infections have been previously reported in 27% - 43% of LGV cases (Spaargaren, Fennema, *et al.*, 2005; Saxon *et al.*, 2016).

Whole genome sequencing (WGS) has provided insight into the genomic diversity of LGV *C. trachomatis* strains. In 2007, the genomes of two LGV strains – L2/434/Bu, which causes the inguinal LGV syndrome and was originally isolated from an inguinal bubo in California; and

L2b/UCH-1/proctitis, which was isolated from a patient with LGV proctitis in London – were sequenced (Thomson *et al.*, 2007). It was found that the L2b genome was nearly identical to that of strain L2/434/Bu. Each genome consisted of a 1.04 Mb chromosome and a 7.5 kb plasmid, encoding 889 and 8 coding sequences (CDS), respectively (Thomson *et al.*, 2007). However, the L2 and L2b strains are differentiated by 573 single nucleotide polymorphisms (SNPs) across their genome. Of these, 457 SNPs occur within CDSs. The highest variation between the genomes is found within *tarp* (Thomson *et al.*, 2007). The *tarp* gene encodes an effector protein (a translocated actin-recruiting phosphoprotein, TarP) that is secreted via the *C. trachomatis*type III secretion system (T3SS). The secretion of this protein induces actin polymerisation within the host cell, which facilitates the uptake of EBs into an inclusion (Borges and Gomes, 2015b) (**1.4**). Mutations within the *tarp* gene account for more than 40 amino acid alterations amongst LGV strains (Borges and Gomes, 2015a). Sequence variation within *tarp* has also been noted amongst L2b *C. trachomatis* strains (Borges and Gomes, 2015a).

A further region of variation between the genomes of L2 and L2b strains is within the polymorphic membrane protein H (*pmpH*) gene. This gene encodes an auto-transporter polymorphic membrane protein (Seth-Smith *et al.*, 2013, 2021). PmpH is located on the elementary body (EB) surface and is thought to function as an adhesin (Seth-Smith *et al.*, 2021). PmpH is strongly immunogenic and has been shown to elicit a strong pro-inflammatory cytokine response (Byrne, 2010). All LGV strains contain an LGV-specific 36 bp deletion in the *pmpH* gene, and this deletion is exploited in a real-time PCR assay used to diagnose LGV infections (Morré *et al.*, 2005). L2b *C. trachomatis* strains have a characteristic 9 bp insertion in the *pmpH* gene. This insertion was initially thought to be specific to L2b strains, however the insertion was later identified in strains with *ompA* genotype L2 (Touati *et al.*, 2016).

The *ompA* gene is recognised as a mutational hotspot within the *C. trachomatis* genome (Nunes *et al.*, 2009). Amongst LGV strains, there is high variability of *ompA* (Borges and Gomes, 2015a). Since 2012, there has been a resurgence in LGV cases caused by strains with *ompA* genotype L2 (Stary *et al.*, 2008; Peuchant *et al.*, 2016; Isaksson *et al.*, 2017; Cole *et al.*, 2020a) (**Chapter 5**). Novel L2 and L2b *ompA* variants have been detected across Europe (Stary *et al.*, 2008; Cole *et al.*, 2020b; Marangoni *et al.*, 2021; Seth-Smith *et al.*, 2021). High levels of recombination within the *ompA* gene have been evidenced (Somboonna *et al.*, 2011; Borges *et al.*, 2019). One recombination event resulted in an LGV strain with a chimeric genome (Somboonna *et al.*, 2011). Genomic analysis revealed that this recombinant strain contained clustered regions of genetic exchange between L2 and D *C. trachomatis* strains, resulting in a hypervirulent disease phenotype (Somboonna *et al.*, 2011). The L2/D recombinant strain ('L2c') was obtained from the rectum of a HIV-negative MSM

with severe haemorrhagic proctitis. The study suggested that disease severity could be influenced by key genetic variation within the LGV genome.

Genomic analyses of LGV strains have been conducted on strains isolated from patients with symptomatic infections (i.e. the inguinal LGV syndrome or LGV proctitis). Strains isolated from patients with asymptomatic infections have never been analysed. It is unknown whether the absence of disease symptoms is the result of genetic variation within the genome, or a host effect. Further investigation into the genetic characteristics of asymptomatic LGV infections may reveal key marker sequences within the genomes.

In this Chapter, LGV isolates collected from asymptomatic and symptomatic MSM in Brighton (Labiran *et al.*, 2016) were whole genome sequenced. The hypothesis for this study was that sequence differences at genomic loci between isolates from symptomatic and asymptomatic MSM in Brighton could account for the difference in patient symptomology. The resulting genomes were compared in order to investigate genomic diversity within these isolates and to indicate genomic loci that may be linked with the presence (or absence) of clinical symptoms.

6.2 Aims and Objectives

The aim of this Chapter was to investigate genomic diversity within LGV strains collected from symptomatic and asymptomatic MSM in Brighton. We wished to determine whether sequence differences at genomic loci within the genomes of these strains could potentially explain the differences in symptomology.

To achieve this aim, the objectives were as follows:

- To undertake whole genome illumina sequencing of genomic DNA from three LGV isolates obtained from symptomatic and asymptomatic MSM in Brighton.
- To assemble and annotate whole genome sequences for each LGV isolate.
- To compare the three assembled genome sequences and identify regions of interest for future studies.

6.3 Study Design

This comparative genomic study was designed to analyse and compare the genome sequences of LGV strains isolated from symptomatic MSM and those isolated from asymptomatic MSM, in order to determine if any regions of sequence variation within the genomes existed.

A previous study in our research group collected *C. trachomatis*-positive rectal swab specimens from MSM attending the Claude Nichol GUM Clinic in Brighton (Labiran *et al.*, 2016). A number of these swab specimens were isolated using cell culture, including seven LGV isolates; these isolates were subsequently MLVA-*ompA* genotyped (Labiran *et al.*, 2016). The symptomology of each patient providing an isolate was recorded by the clinician. Table 24 shows the symptomology and the MLVA-*ompA* genotypes for each of the 7 cultured LGV isolates.

Of the seven patients providing LGV isolates in the Labiran study, five reported rectal symptoms (Bri038, Bri085, Bri087, Bri088, and Bri104) and two reported no symptoms (Bri086 and Bri098) (**Table 25**). In this Chapter, we specifically wanted to utilise whole genome sequencing to compare the genome sequences of LGV isolates collected from symptomatic and asymptomatic MSM. This method was selected as whole genome sequencing provides the maximum resolution for strain discrimination.

It was initially decided to sequence and compare the genomes of two symptomatic LGV isolates (Bri038 and Bri088) and two asymptomatic LGV isolates (Bri086 and Bri098). These isolates had previously been grown in cell culture and the harvested *C. trachomatis* stored at -80°C (**6.4.1**). Whilst genomic DNA (gDNA) was extracted from the isolates in the earlier study (Labiran *et al.*, 2016) for MLVA-*ompA* genotyping, we could not use these gDNA extracts for whole genome sequencing as the gDNA was of insufficient volume and concentration. For this Chapter, the harvested *C. trachomatis* was thawed and re-grown in cell culture (**6.4.2**). The first isolate to be grown in cell culture was Bri086. This isolate was discovered to be contaminated. When the tissue culture flask containing the Bri086-infected McCoy cell monolayer was visualised under the phase contrast microscope (**2.17**), rod-shaped structures were present, indicating contamination by enteric bacteria such as *E. coli*. Vancomycin (50 µL of 1 mg/ml) and gentamycin (10 µL of 10 mg/ml) were added to treat the *E. coli* contamination. When the flask was titrated, no *C. trachomatis* was present. For Bri038, Bri088, and Bri098, it was decided to add the bacterial inocula (200 µL from Labiran harvested stock), vancomycin, gentamycin and cycloheximide in 5 ml of DMEM to the McCoy cell monolayer (**6.4.2**). When tissue culture flasks containing McCoy cell monolayers infected with Bri038, Bri088 and Bri098 were visualised under the phase-contrast microscope, no *E. coli* contamination was present. All 3 isolates produced good-sized *C. trachomatis* inclusions that were visible after 24 hours (**6.4.2**). *C. trachomatis* was harvested from each of the three flasks after 48 hours. Genomic DNA was extracted (**6.4.2**) and sent for whole genome sequencing (**6.4.2**).

Table 25 Cultured rectal Isolates collected from MSM[1] attending the Claude Nichol GUM[2] Clinic in Brighton between 2011 and 2013. Isolates were collected using rectal swabs, grown in cell culture, and MLVA-*ompA* genotyped by Labiran *et al.* (2016). Symptomology data was available for each patient.

Isolates	**Symptoms (Y/N)***	**MLVA-*ompA* genotype**
Bri038	Y	1.9.2 – L2b
Bri085	Y	1.5.2 – L2b
Bri086	N	1.9.2 – L2b
Bri087	Y	1.5.3 – L2b
Bri088	Y	1.9.2 – L2b
Bri098	N	3.5.2 – L2b
Bri104	Y	1.9.4 - L2

[1] MSM = men who have sex with men; [2] GUM = Genitourinary Medicine Clinic; *Symptoms reported as "Y" refer to rectal symptoms.

6.4 Methods

6.4.1 C. trachomatis isolates

The three isolates that were whole genome sequenced in this Chapter (Bri038, Bri088, and Bri098) were originally sourced from MSM attending the Claude Nichol Clinic in Brighton between 2011 and 2013 (Labiran *et al.*, 2016). Isolates were collected from patients using rectal swabs, and these swabs were immediately placed in Σ-VCM™ Universal Transport Medium (*Sigma-Aldrich, UK*). This Medium is specially formulated to allow the survival and recovery of Viruses, Chlamydia, Mycoplasma and Ureaplasma from swabs. According to Labiran (2014), the swab transport medium for each isolate was vortexed followed by centrifugation at 110 xg for 5 minutes, and was subsequently diluted 1:10 using DMEM, before placing on a McCoy cell monolayer. McCoy cells infected with each isolate (**2.16**) were incubated for 48 hours after which *C. trachomatis* was harvested (**2.17**). The cryopreservative 4X Sucrose Phosphate (4SP) was added and harvested samples were subsequently stored at -80°C. Labiran extracted genomic DNA from each harvested sample using the Wizard® Genomic DNA Purification Kit (*Promega, UK*).

The Labiran *et al*. (2016) study also collected data relating to symptomology from each patient. Bri038 and Bri088 were collected from symptomatic MSM (MSM with rectal symptoms), and Bri098 was collected from an asymptomatic MSM (**Table 25**).

6.4.2 Revival of *C. trachomatis* LGV isolates from Brighton in 2020

Each harvested isolate was removed from the -80°C freezer and thawed. For each isolate, 5 ml fresh DMEM containing bacterial inocula, cycloheximide (1 µg/ml), vancomycin (10 µg/ml), and gentamycin (20 µg/ml) was added to a McCoy cell monolayer (**2.14.2; 2.16**). After 24 hours, *C. trachomatis* inclusions were visible. *C. trachomatis* isolates were harvested 48 hours post-infection (**2.17**). Phase contrast microscope images showing the McCoy cell monolayers infected with each of the three isolates were taken at 48 hours post-infection (**Figure 23**). Inclusions are visible within the McCoy cell monolayer of each tissue culture flask (one inclusion per flask is indicated with a black arrow).Genomic DNA was extracted from each using the NucleoSpin® Tissue kit (*Macherey-Nagel, Germany*) (**2.4.2**).

The MLVA-*ompA* genotype of Bri038 was determined prior to sending the genomic DNA for whole genome sequencing, to confirm that the genotype was the same as that obtained by Labiran (2014). This process was described in Chapter 4 (4.4.1). The MLVA-*ompA* genotype of Bri038 was identified as 1.9.2b - L2b. This was the same MLVA-*ompA* genotype as reported by Labiran (2014). To note, the VNTR sequences and *ompA* sequence were identical to those obtained from Bri038 by Labiran, however Labiran noted the CT1291 VNTR type as 2 and not 2b (**4.4.1**). It was deemed unnecessary to MLVA-*ompA* genotype the remaining isolates given that we could later confirm the MLVA-*ompA* genotype using the whole genome sequence.

6.4.3 Whole genome sequencing

Genomic DNA (150 ng per isolate) was sent for whole genome illumina sequencing at MicrobesNG (**2.18**). MicrobesNG used 250 bp paired-end reads, and adapter trimmed using Trimmomatic (version 0.30) with a sliding window quality cut-off of Q15 (Bolger, Lohse and Usadel, 2014). MicrobesNG performed de novo assembly on each sample using SPAdes (version 3.7) (Bankevich *et al.*, 2012), and annotated contigs using Prokka (version 1.11) (Seemann, 2014). The number of reads, number of contigs, mean coverage (average depth of reads across the genome), and N50 for each of the three isolates can be found in Table 26.

Table 26 The number of reads, number of contigs, mean coverage, and N50* values for the three rectal isolates, Bri038, Bri088, and Bri098.

	Number of reads	Number of contigs	Mean Coverage (average read depth)	N50*
Bri038	137,813	282	23.2X	1,033,306
Bri088	216,959	2,963	5.4X	792
Bri098	148,953	97	46.2X	1,038,696

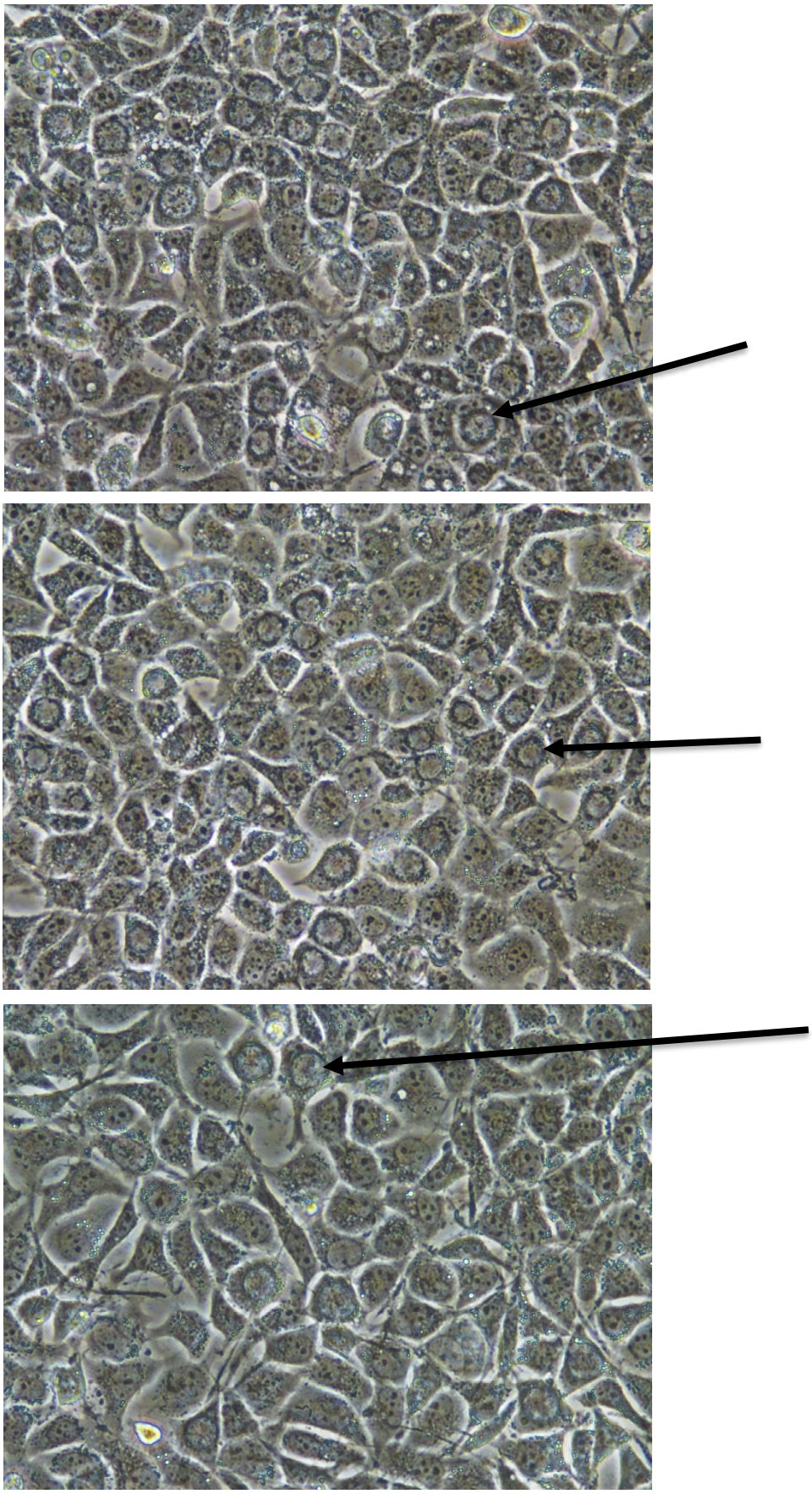
*N50 represents the smallest contig such that half of the genome is represented by contigs of size N50 or larger).

Contigs were re-annotated using Prokka with the *Chlamydia* genus and a list of proteins (in fasta format) derived from the reference genome L2b/UCH-1/proctitis (Genbank Accession No. AM884177) as a basis for annotation. The Galaxy server was used to perform these analyses (Afgan et al., 2018).

6.4.4 Assignation of MLST and MST sequence types (STs) for Bri038, Bri088, and Bri098

MLST (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) and MST (Klint *et al.*, 2007) sequence types (STs) were assigned to the 3 isolates using the method described in section 4.4.3.

(A)



(B)

(C)

Figure 23 Phase Contrast microscope images of McCoy cell monolayers infected with three LGV *C. trachomatis* isolates in a T25 cell culture flask. (A) Bri038; (B) Bri088; (C) Bri098. Images were taken 48 hours post-infection using x40 magnification. Inclusions (indicated with a black arrow) are visible within McCoy cells in all three flasks.

6.5 Genomic Strain Comparison

To identify genomic differences between the three LGV isolates, sequencing reads (fastq files) for each isolate were mapped to the L2b/UCH-1/proctitis reference genome sequence (Genbank Accession No. AM884177) using Snippy (version 2.2) (Seemann, 2018). Read data was also mapped against the plasmid sequence for L2b/UCH-1/proctitis (Genbank Accession No. AM886279). This identified any single nucleotide polymorphisms (SNPs), insertions (ins), or deletions (del) within each genome (**Table 27**). Snippy was accessed using the Galaxy server (Afgan et al., 2018). The raw Snippy output data can be found in Appendix 9. All called SNPs were checked manually using BAM files visualised in Artemis.

A total of eleven mutations were identified between the three isolates relative to L2b/UCH-1/proctitis. All eleven mutations were located within the chromosome; no mutations were detected within the plasmid of each isolate. The plasmid sequence of each isolate was identical to that of L2b/UCH-1/proctitis (Genbank Accession No. AM886279). No mutations were shared by all 3 isolates. Of the 11 mutations detected, 1 was a deletion, 2 were insertions, and 8 were single nucleotide polymorphisms (**Table 27**). The deletion was observed within an intergenic region of Bri098 only. The two insertions were also within intergenic regions, however one insertion was observed in Bri088 only, and the second insertion was observed in both Bri088 and Bri098.

Of the eight SNPs identified, six were missense mutations within coding sequences (CDS) that resulted in amino acid substitutions in the corresponding proteins (**Table 27**). Two SNPs (25%) were within the *ompA* gene encoding the major outer membrane protein (MOMP). One SNP at nucleotide position 59,310 (nucleotide position given relative to L2b/UCH-1/proctitis) caused a change from leucine to isoleucine in the amino acid sequence of the MOMP (L173I). This SNP was observed in both Bri038 and Bri088, but was not observed in Bri098. The second SNP (at nucleotide position 59,342) in *ompA* was present in Bri088 and Bri098, and resulted in the amino acid serine being substituted by asparagine (S162N). This substitution altered the *ompA* genotype of Bri088 and Bri098 from L2b to L2. These two non-synonymous mutations within *ompA* occurred within variable segment 2 of the *ompA* gene, which encodes variable domain 2 of the MOMP (**Figure 24**).

The three remaining missense mutations identified were each observed in one isolate only. In Bri088, a SNP at nucleotide position 615,012 was present in the *accA* gene encoding Acetyl-coenzyme A (CoA) carboxylase carboxyl transferase subunit alpha. This SNP resulted in the amino acid tyrosine being substituted by serine (Y284S). This SNP in *accA* was not observed in Bri038 or Bri098. In Bri038, a missense mutation (at nucleotide position 630,873) in the *nqrE* gene that encodes the Na^+-translocating NADH-quinone reductase subunit E caused an amino acid change from glycine to aspartic acid (G74D). The two remaining non-synonymous mutations were observed

in Bri098 only, and resulted in amino acid substitutions within two hypothetical proteins (**Table 27**). These were at nucleotide positions 72,990 and 760,790, and resulted in the amino acid changes Q124K and V541I, respectively.

Table 27 Mutations in Bri038, Bri088, and Bri098, relative to L2b/UCH-1/proctitis (Genbank Accession No. AM884177). *these SNPs were located on the reverse strand and have been reverse complemented.

Isolates	Type	Location	Effect	Gene Product
Bri038 Bri088	snp (C->A)*	CDS	Missense (L->I)	Major outer membrane protein (MOMP)
Bri088 Bri098	snp (G->A)*	CDS	Missense (S->N)	Major outer membrane protein (MOMP)
Bri098	snp (C->A)	CDS	Missense (Q->K)	Hypothetical protein
Bri098	snp (A->G)	Intergenic		
Bri088	ins (T->TG)	Intergenic		
Bri038 Bri088	snp (G->A)	Intergenic		
Bri088	snp (A->C)*	CDS	Missense (Y->S)	Acetyl-coenzyme A (CoA) carboxylase carboxyl transferase subunit alpha
Bri038	snp (G->A)	CDS	Missense (G->D)	Na (+)-translocating NADH-quinone reductase subunit E
Bri098	snp (G->A)	CDS	Missense (V->I)	Hypothetical protein
Bri098	del (AT->A)	Intergenic		
Bri088 Bri098	ins (C->CT)	Intergenic		

Ins = insertion; snp = single nucleotide polymorphism; del = deletion; CDS = coding sequence.

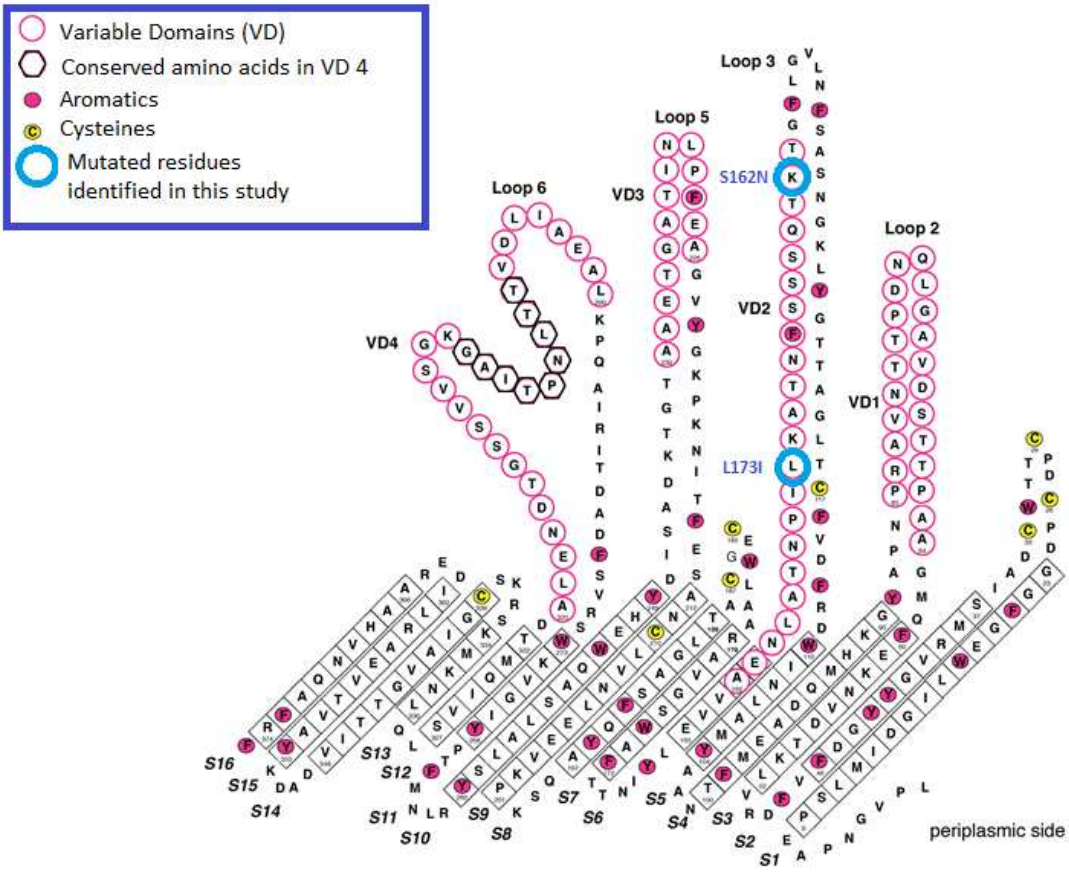


Figure 24 Location of altered amino acids (circled in blue) within the major outer membrane protein (MOMP) of Bri038, Bri088, and Bri098.

Figure is adapted from the model developed in Fehrer *et al*., (2013) that was based on the *C. trachomatis ompA* genotype C sequence (Genbank Accession No. DQ116399). Amino acid substitutions in Bri038, Bri088, and Bri098 differing from those in L2b/UCH-1/proctitis are given next to the position of the amino acid (positions given relative to L2b/UCH-1/proctitis). Bri038 and Bri088 contained the SNP* resulting in the amino acid change L173I, and Bri088 and Bri098 contained the SNP* resulting in the amino acid change S162N. The two amino acid substitutions are located within variable domain 2 (VD 2) of the MOMP.

*SNP = single nucleotide polymorphism.

None of the mutations identified in Bri038, Bri088, and Bri098 were located within the genetic loci comprising the MLST (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) and MST (Klint *et al.*, 2007) schemes for *C. trachomatis*. The MLST and MST allelic profiles and sequence types (STs) for Bri038, Bri088, and Bri098 are provided in Table 26. All 3 isolates shared the same STs – ST 44 (Pannekoek *et al.*, 2008), ST 1 (Dean *et al.*, 2009), and ST 58 (Klint *et al.*, 2007). These STs were identical to those of L2b/UCH-1/proctitis (Genbank Accession No. AM884177) (**Table 28**).

Table 28 MLST (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) and MST (Klint *et al.*, 2007) sequence types (STs) for Bri038, Bri088, and Bri098. The MLST and MST STs for the L2b reference genome L2b/UCH-1/proctitis (Genbank Accession No. AM884177) are also provided in the table.

	MLST (Pannekoek *et al.*, 2008)		**MLST (Dean *et al.*, 2009)**		**MST (Klint *et al.*, 2007)**	
	Allelic profile[1]	**ST**	**Allelic Profile[2]**	**ST**	**Allelic Profile[3]**	**ST**
Bri038	1-3-3-3-2-2-19	44	1-1-3-8-1-4-11	1	13-17-13-27-28	58
Bri088	1-3-3-3-2-2-19	44	1-1-3-8-1-4-11	1	13-17-13-27-28	58
Bri098	1-3-3-3-2-2-19	44	1-1-3-8-1-4-11	1	13-17-13-27-28	58
L2b/UCH-1/proctitis	1-3-3-3-2-2-19	44	1-1-3-8-1-4-11	1	13-17-13-27-28	58

[1] For MLST (Pannekoek *et al.,* 2008), allelic profiles given in the order: *enoA, fumC, gatA, gidA, hemN, hlfX*, and *oppA*.
[2] For MLST (Dean *et al.,* 2009), allelic profiles given in the order: *glyA, mdhC, pdhA, yhbG, pykF, lysS*, and *leuS*.
[3] For MST (Klint *et al.,* 2007), allelic profiles given in the order: *CT058, CT144, CT172, pbpB, and hctB.*

The VNTR types of Bri038, Bri088, and Bri098 were 1.9.2b (CT1335 type 1: GAAAAAG-**9T8A**-GCTTTTGT; CT1299 type 9: TTTTTATTCT-3C2T-**6C**-ATCAAA; CT1291 type 2b: AAAATAGTCTA-**8C**-TATTG). This VNTR type was identical to that of L2b/UCH-1/proctitis (**Chapter 4**). The *ompA* genotypes of the isolates were L2b (Bri038) and L2 (Bri088 and Bri098). The MLVA-*ompA* genotype assigned by Labiran (2014) to Bri098 was 3.5.2- L2b. Sequence chromatograms generated by Labiran (2014) were viewed to compare the sequence at each VNTR locus to those obtained in this Chapter by whole genome sequencing. The sequence quality obtained by Labiran at each VNTR locus for Bri098 was poor hence the Labiran assignation of the genotype 3.5.2 should be treated with caution (**Appendix 10**). The 1,185 bp Bri098 *ompA* sequence that was obtained by whole

genome sequencing in this Chapter was aligned against the 923 bp *ompA* sequence obtained for Bri098 by Labiran (2014) (**Appendix 11**). These sequences were identical. The SNP that reverts the L2b *ompA* sequence to L2 (G485A) was present in both sequences, confirming that Bri098 was *ompA* genotype L2 (**Appendix 11**).

The *pmpH* gene sequence of the three LGV isolates were identical to L2b/UCH-1/proctitis. Each of the isolates contained the characteristic 36 bp deletion within the *pmpH* gene, in addition to the 9 bp insertion that has previously been detected within L2b and L2 *C. trachomatis* strains (**Figure 25**).

To observe the relative genetic relatedness of Bri038, Bri088, Bri098 and L2b/UCH-1/proctitis, the data generated in Table 25 was combined to produce a core SNP genome using Snippy core (version 4.6.0). MEGA-X (version 10.1.8) was used to produce a phylogenetic tree based on the divergence of each isolate from the core genome, using the Generalised Time Reversible (GTR) model with 1,000 bootstraps. The phylogenetic tree was rooted on L2b/UCH-1/proctitis (**Figure 26**).

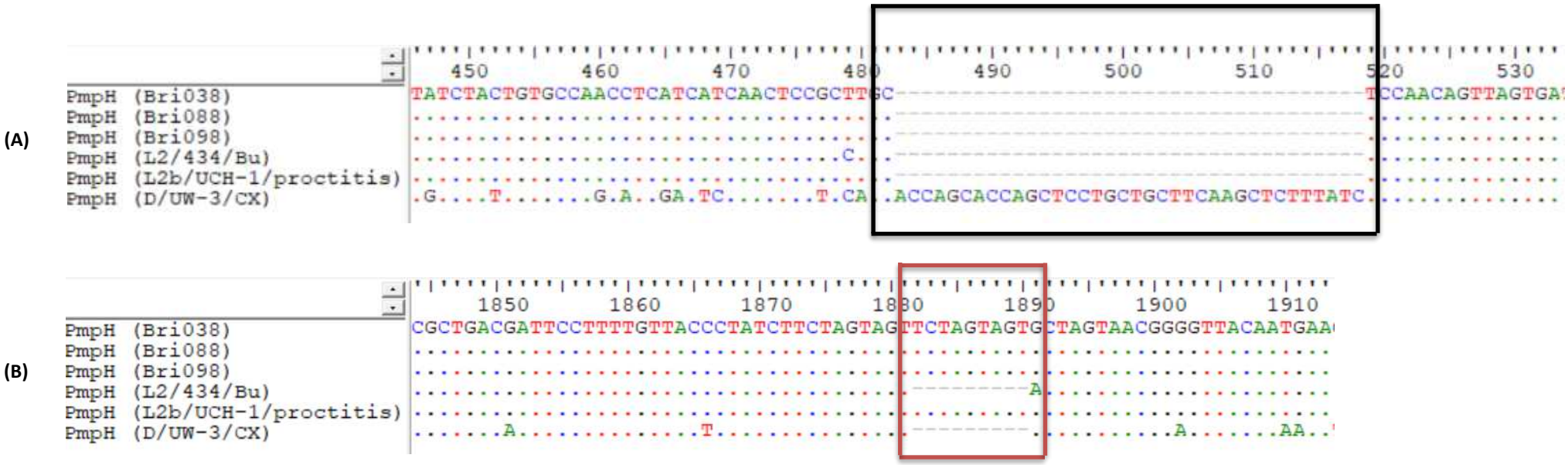


Figure 25 Sequence Alignment of sections of the *pmpH* gene from three LGV isolates from Brighton. The *pmpH* sequences were obtained from the three isolates by whole genome sequencing. Also included in the sequence alignment were the *pmpH* sequences of L2/434/Bu (Genbank Accession No. AM884176), L2b/UCH-1/proctitis (Genbank Accession No. AM884177), and D/UW-3/CX (Genbank Accession No. NC_000117) for comparison. (A) shows the characteristic 36 bp deletion in the *pmpH* gene of LGV *C. trachomatis* strains (indicated by a black box). (B) shows the 9 bp insertion in the *pmpH* gene of L2b *C. trachomatis* strains (indicated by a red box). This insertion is not present in the L2/434/Bu strain, however the insertion has been identified in other L2 strains (Touati *et al.*, 2016).

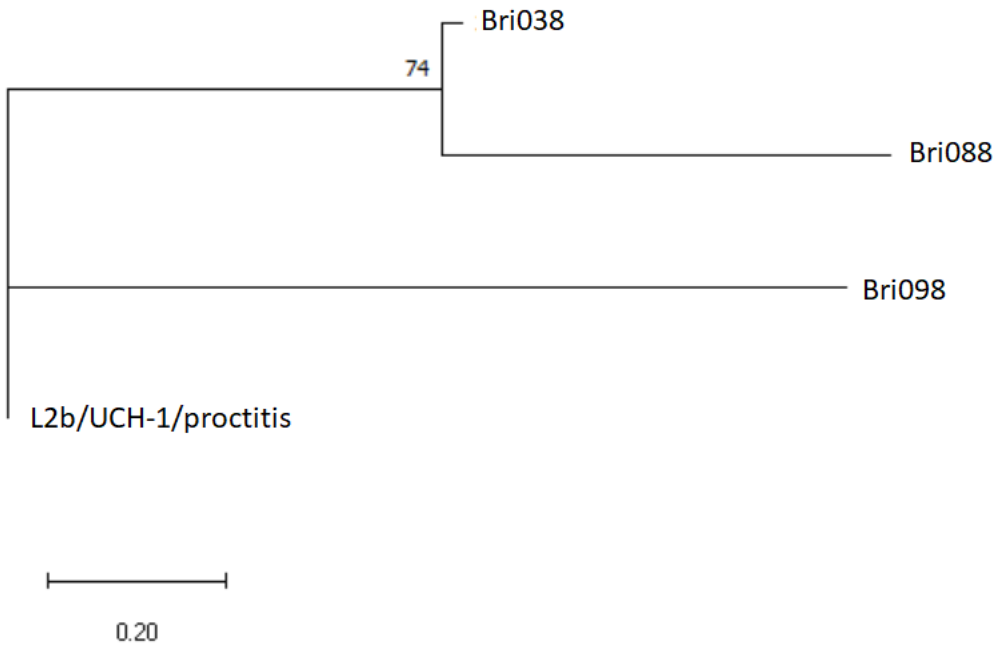


Figure 26 Phylogenetic Tree of Bri038, Bri088, Bri098, and L2b/UCH-1/proctitis.

Sequencing reads for isolates Bri038, Bri088, and Bri098 were mapped to the L2b reference genome, L2b/UCH-1/proctitis (Genbank Accession No. AM884177) using Snippy (version 2.2). A core SNP genome was produced using Snippy core (version 4.6.0). MEGA-X (version 10.1.8) was used to produce a phylogenetic tree based on the divergence of each isolate from the core SNP genome, using the Generalised Time Reversible (GTR) substitution model and 1,000 bootstrap replications. The phylogenetic tree was rooted on L2b/UCH-1/proctitis. The branch lengths indicate the number of substitutions per variant site.

6.6 Discussion

The incidence of LGV has increased in recent years (Public Health England, 2020d). The proportion of asymptomatic LGV infections has also increased (Cole *et al.*, 2020b). In the Netherlands, the proportion of asymptomatic LGV infections has increased from 31.4% in 2011 to 49.3% in 2017 (van Aar *et al.*, 2020). No comparative genomic analysis has ever been conducted between LGV isolates obtained from patients with asymptomatic LGV infections, and LGV isolates obtained from patients with symptomatic LGV infections.

A study by Somboonna *et al.* (2011) analysed the genome of an LGV strain with a hypervirulent disease phenotype that produced severe haemorrhagic proctitis in a HIV-negative MSM. This study identified that the strain possessed a chimeric L2/D genome and was a recombinant of L2 and D *C. trachomatis* strains. There were extensive clustered regions of genetic exchange including a 78 kb region and a partial toxin gene within the strain genome. Indels (insertions/deletions) were identified within an *ftsK* gene promoter, and in the *tarp* and *hctB* genes. These regions encode proteins involved in replication, inclusion formation, and histone H1-like protein activity, respectively. The study concluded that the significant disease pathology of the strain may be attributed to its widespread genetic recombination (Somboonna *et al.*, 2011).

In this Chapter, we utilised whole genome sequencing to sequence three LGV isolates collected from asymptomatic and symptomatic MSM in Brighton. Two of these isolates (Bri038 and Bri088) were obtained from MSM with symptomatic LGV infections, and one isolate (Bri098) was collected from an MSM with an asymptomatic LGV infection. We hypothesised that sequence differences at genomic loci between isolates from symptomatic and asymptomatic MSM could account for the difference in patient symptomology. The genomes were compared to investigate genomic diversity within the isolates and to identify genomic loci that could be linked with the presence or absence of clinical symptoms. Studies of genotype-phenotype using a whole genome sequencing approach have previously been conducted in other species within the *Chlamydia* genus. As discussed in section 1.5.2, Ramsey *et al.* (2009) identified several unique non-synonymous SNPs when comparing two *C. muridarum* strains (*C. muridarum* Weiss and *C. muridarum* Nigg) that varied in virulence in vivo and possessed distinct growth characteristics in vitro. A limitation of this study was that it was unclear whether the isolates used in this study were representative of the entire population. However the study did show that genomic differences exist within stock isolates that varied in virulence. Similarly, genotype-phenotype studies that have adopted a WGS approach have also been conducted in *C. trachomatis*. A study by Kari *et al.* (2008) analysed the genomes of four trachoma strains to assess whether sequence variation played a role in disease outcome. The study identified a subset of genes that were polymorphic amongst the four trachoma strains and

demonstrated that most of these polymorphic genes also varied in recent clinical trachoma isolates. It was shown that six genes were likely associated with distinct phenotypic properties within the trachoma strains, including growth rate, such as *hctB*.

In this Chapter, we identified 11 mutations within the LGV isolates compared to the genome of the L2b proctitis outbreak strain, L2b/UCH-1/proctitis (**Table 26**). This low number of SNPs distinguishing the isolate genomes from that of L2b/UCH-1/proctitis indicates that the isolates have an L2b genomic backbone. In comparison, the genome of *C. trachomatis* L2 strains differ from L2b/UCH-1/proctitis by 573 SNPs (Thomson *et al.*, 2007).

Two of the 11 mutations identified between the three isolates and L2b/UCH-1/proctitis were within the *ompA* gene encoding the major outer membrane protein (MOMP). Both mutations were non-synonymous SNPs that resulted in amino acid changes in variable domain 2 of the MOMP (L173I and S162N) (**Figure 24**). The variable domains encode surface-exposed epitopes, and amino acid changes in these domains may affect the epitopes targeted by the hosts' immune system. A key finding of the study was that Bri088 and Bri098 contained a SNP within *ompA* that reverted the L2b *ompA* genotype to L2 (G485A resulting in S162N) (**Table 26**).The isolate Bri038 did not contain this SNP. Our results indicate that Bri088 and Bri098 have an L2b genomic backbone and an L2 *ompA* genotype, whilst Bri038 has an L2b genomic backbone and an L2b *ompA* genotype. The reversion of the L2b genomic backbone to *ompA* genotype L2 has previously been reported by Seth-Smith *et al.* (2021). The Seth-Smith *et al.* (2021) study utilised whole genome sequencing to sequence 42 LGV-positive specimens, and identified 26 genomes with *ompA* genotype L2 and an L2b genomic backbone, and three genomes with *ompA* genotype L2b and an L2b genomic backbone. All 26 genomes with *ompA* genotype L2 contained the SNP reversion identified in our study at position 59,342 (G485A, resulting in the amino acid change S162N). It has been suggested that the reversion to *ompA* genotype L2 may be due to strong selective pressure on the immunogenic variable domains of MOMP, such that chance mutations within the variable domains are selected for in circulating strains (Seth-Smith *et al.*, 2021). The L2 *ompA* genotype is predicted to be slightly less antigenic than that of L2b (Seth-Smith *et al.*, 2021), hence infections with strains with *ompA* genotype L2 may be at a selective advantage within a naïve host. This was determined through modelling of MOMP epitope antigenicity, which indicated that L2b MOMP with a serine at position 162 had a higher antigenicity than L2 with a glutamine at position 162 (Seth-Smith *et al.*, 2021). However, it should be noted that the antigenicity scores for L2b (0.939) were only marginally higher than that of L2 (0.906), hence caution must be applied when interpreting and forming conclusions from these results. In the London population, we observed the L2b genomic backbone in all 3 LGV isolates in this study, including the two isolates with *ompA* genotype L2 (Bri088 and Bri098). These results are consistent with the findings of Harris *et al.* (2012) and Hadfield *et al.* (2017) that

determined that the LGV proctitis outbreak is the result of clonal expansion of L2b within high-risk sexual networks. It was suggested by Seth-Smith *et al.* (2021) that the combination of the L2b genomic backbone with *ompA* genotype L2 may improve the chances of infection in populations with pre-existing immunity to *ompA* genotype L2b. However, owing to the small number of isolates included in our study, we are unable to form any conclusions relating to this.

No mutations were present within the *pmpH* gene of the three isolates, and the *pmpH* gene sequence was identical to that of L2b/UCH-1/proctitis. Our results show that all three isolates contained the 36 bp deletion that is characteristic of all LGV strains, in addition to the 9 bp insertion that has previously been detected amongst L2b and L2 strains of *C. trachomatis* (Touati *et al.*, 2016) (**Figure 25**). The 36 bp deletion within the *pmpH* gene is targeted by a real-time PCR assay used to detect LGV in clinical specimens. Our results demonstrate that this assay can still detect LGV within isolates with an L2b genomic backbone and an L2 *ompA* genotype.

The eleven mutations identified within the 3 isolates compared to L2b/UCH-1/proctitis were located within the chromosome; the three isolates had an identical plasmid sequence to that of L2b/UCH-1/proctitis. None of the 11 mutations identified within the 3 isolates from Brighton were within the genetic loci targeted by the MLST (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) and MST (Klint *et al.*, 2007) genotyping systems. All three isolates were assigned ST 44 (Pannekoek *et al.*, 2008), ST 1 (Dean *et al.*, 2009), and ST 58 (Klint *et al.*, 2007) (**Table 27**). These STs were identical to those assigned to L2b/UCH-1/proctitis (**Table 27**). The MLVA-*ompA* VNTR type (Pedersen *et al.*, 2008) of all three isolates was 1.9.2b. This VNTR type was identical to that assigned to L2b/UCH-1/proctitis. In this study, Bri098 was assigned the complete MLVA-*ompA* genotype 1.9.2b - L2. This contrasted with the MLVA-*ompA* genotype previously assigned to this isolate by Labiran (Labiran *et al.*, 2016), which was 3.5.2 - L2b. The sequence chromatograms for the three VNTR loci obtained by Labiran were viewed to compare against the sequences obtained by whole genome sequencing. The sequence quality of the chromatograms was poor and contained a high level of background (**Appendix 10**), making it difficult to determine the VNTR type at each locus from the sequence traces. Furthermore, Labiran assigned Bri098 the *ompA* genotype L2b, whereas in our study it was assigned the *ompA* genotype L2. The *ompA* sequence obtained for Bri098 in our study was aligned against the *ompA* of Bri098 obtained by Labiran (**Appendix 11**). The sequence alignment showed that the *ompA* sequences obtained in both studies were identical. Both sequences showed the SNP reversion of L2b to L2 (G485A, resulting in the amino acid change S162N) in Bri098. Given these findings, we can conclude that the *ompA* genotype L2b was mis-attributed to Bri098 by Labiran (Labiran *et al.*, 2016). This confirmed the assignation of *ompA* genotype L2 to Bri098 in our study.

We identified that 75% of the SNPs in protein coding regions across the genome are non-synonymous. This supports previous observations in the L2b clade of 75% - 90.2% of the SNPs in CDSs being non-synonymous (Borges and Gomes, 2015a; Seth-Smith *et al.*, 2021). Outside of the *ompA* gene, there were 4 non-synonymous SNPs in CDSs compared to L2b/UCH-1/proctitis. The two isolates obtained from symptomatic MSM (Bri038 and Bri088) contained SNPs within genes that are thought to play functional roles in intracellular replication (*nqrE* and *accA*) and transition of EBs to RBs (*accA*). Bri038 contained a non-synonymous SNP within the *nqrE* gene. This gene encodes the Na^+- translocating NADH – quinone reductase subunit E. This subunit is one of six subunits that comprise the sodium ion (Na^+) - pumping NADH-quinone oxidoreductase (Na^+ - NQR) enzyme. This enzyme acts as a sodium ion pump and generates a sodium ion gradient that drives metabolic processes within *C. trachomatis*. A study by Liang *et al.* (2018) showed that the enzyme sustains aerobic metabolism that is required for RB replication. The effect of the non-synonymous SNP within *nqrE* on the Na^+ - NQR enzyme in Bri038 is unclear, however the SNP was not present in Bri088 or Bri098. Bri088 (also isolated from a symptomatic MSM) contained a non-synonymous mutation within the acetyl co-enzyme A (CoA) carboxylase subunit alpha (*accA*) gene. This gene encodes one of the four subunits that comprise the acetyl CoA carboxylase enzyme complex (AccABCD). The AccABCD enzyme complex catalyses the conversion of acetyl CoA to malonyl CoA in the first step of fatty acid synthesis (Yao and Rock, 2018). Fatty acid synthesis is required for the transition of EBs into RBs as well as the replication of RBs (Yao and Rock, 2018). The SNP within *accA* of Bri088 is not present in either Bri038 or Bri098. It is of interest that the two isolates collected from symptomatic MSM (and not the isolate collected from an asymptomatic MSM) had non-synonymous SNPs within genes with roles in *C. trachomatis* infection and growth. The asymptomatic isolate, Bri098, contained two SNPs within hypothetical proteins compared to L2b/UCH-1/proctitis. These SNPs were not present within the two isolates obtained from symptomatic MSM. All of these SNPs were non-synonymous, which may suggest that there is a selection pressure within these genes. It is unclear whether mutations within these genes contribute to the difference in symptomology that was exhibited by patients infected with Bri038, Bri088, and Bri098. Future studies could apply whole genome sequencing to a larger sample size of symptomatic and asymptomatic isolates in order to further investigate causality between the genome sequence and presence or absence of clinical symptoms. Whilst outside of the scope of this study, other factors including the host immune response to *C. trachomatis* may also impact clinical manifestation and should be further investigated. In this way, a limitation with taking solely a WGS approach to considering clinical outcome is that it is a one-sided approach to a complex, multifactorial issue. It considers only one factor – that sequence variation within the pathogen accounts for differences in clinical manifestation in the host. Further, whilst we have identified polymorphisms within the genomes of these three LGV isolates, we cannot form any conclusions

about whether this genetic variation contributed to the symptomatic/asymptomatic phenotype in the host. In this way, it would perhaps have been more appropriate to combine comparative WGS analysis with in vitro or in vivo experimentation, in order to experimentally test whether the subtle genetic differences identified resulted in different pathogenic properties in the three LGV isolates. These methods were adopted by Kari *et al.* (2009), Ramsey *et al.* (2009), and Miyairi *et al.* (2011) (these are discussed further on pages 13-14 and 160). Specifically, Kari *et al.* (2009) was able to show that genetic variations manifested in differences in virulence between two trachoma-causing strains, through measurement of factors such as in vitro growth rate, sensitivity to IFN-γ, and infectivity of both strains in a non-human primate model.

6.7 Conclusions

The study has shown that genomic diversity is present within LGV isolates collected from asymptomatic and symptomatic MSM. We have shown that all three isolates possessed an L2b genomic backbone, however there was variation within the *ompA* gene sequence of the three isolates. Bri088 and Bri098 contained a SNP that reverted the *ompA* genotype from L2b to L2, whereas Bri038 was assigned *ompA* genotype L2b. The study identified that the two isolates collected from MSM with symptomatic LGV infections contained SNPs within the coding sequences of genes associated with intracellular multiplication of RBs, and that these SNPs were absent from the isolate collected from an asymptomatic MSM (Bri098). Further research combining in vitro models and comparative WGS analysis for a larger number of asymptomatic and symptomatic LGV isolates is required in order to more fully investigate these differences in clinical outcome within patients with LGV infections.

CHAPTER 7 Final Discussion

The overall strategic aim of this PhD was to characterise genetic diversity within *C. trachomatis* strains in high-risk sexual networks. We applied several methods including *ompA* genotyping, *trpA* gene sequencing, MLVA-*ompA* genotyping, and whole genome sequencing, in order to achieve this strategic aim.

In Chapter 3, we determined the *ompA* genotype of *C. trachomatis* DNA extracted from a conjunctival swab collected from a 10 year old female. The child had recently been resident in Afghanistan and had presented in clinic with symptoms of conjunctivitis. A positive *C. trachomatis* diagnosis in children outside of the neonatal period is un-common outside of trachoma-endemic countries, and is indicative of child sexual abuse (Hammerschlag, 1998). We utilised *ompA* genotyping to ascertain whether the child was infected with a *C. trachomatis* strain causing trachoma (i.e. strains with *ompA* genotypes A – C), or a *C. trachomatis* strain associated with sexually transmitted urogenital tract infections and sexually-acquired adult inclusion conjunctivitis (i.e. strains with *ompA* genotypes D – K), which could have implications for child safe-guarding. Our *ompA* gene sequencing results indicated that the child was infected with a *C. trachomatis* strain with *ompA* genotype C. Strains with this *ompA* genotype are exclusively associated with trachoma, and there have been no reports of strains with this genotype causing urogenital disease (Andersson *et al.*, 2016; Hadfield *et al.*, 2017). We confirmed these results by sequencing the tryptophan synthase alpha subunit (*trpA*) gene, as polymorphisms within this gene have been previously shown to differentiate between genital and trachoma-causing *C. trachomatis* strains (Caldwell *et al.*, 2003). PCR amplification of the *trpA* gene yielded a poor quality sequence, and repeat sequencing and repeat PCR failed to improve the sequence quality. We developed a cloning strategy using a pUC19 cloning vector that would allow a full-length *trpA* sequence to be obtained from the *C. trachomatis* DNA. A *trpA* PCR product was ligated into a pUC19 vector that was subsequently transformed into competent *E. coli* cells. The recombinant plasmids were sequenced from six transformed *E. coli* colonies, and *trpA* gene sequences were obtained from each plasmid. Previous studies (Fehlner-Gardiner *et al.*, 2002; Caldwell *et al.*, 2003) have shown that trachoma-causing (and not genital) *C. trachomatis* strains have a triplet deletion at nucleotide position 408 – 410 and a nucleotide deletion at position 531; whereas urogenital strains have two point mutations at position 531 that result in amino acid substitutions that vary amongst the serovars. Our results were inconclusive, and 3 *trpA* sequences contained the mutations indicative of trachoma-causing strains, and the remaining 3 *trpA* sequences contained mutations indicative of genital *C. trachomatis* strains. As a result, we were unable to confirm the nature of the *ompA* genotype C strain through our *trpA* sequence analyses.

Multi-locus genotyping systems including MLVA, MST, and MLST have been developed for *C. trachomatis*. We evaluated four high-resolution genotyping systems in Chapter 4, in order to identify which typing system would be most well-suited to genotype LGV strains in a genotyping survey of LGV in London (**Chapter 5**). This was essentially an in silico genotyping survey of LGV strains, and was carried out using whole genome sequence data obtained from NCBI Genome. We determined the discriminatory capacity and typeability for each system. Our results showed that the two MLST schemes (Pannekoek *et al.*, 2008; Dean *et al.*, 2009) had the lowest resolution of the four typing systems evaluated. This was partly expected as these two schemes were based on stable housekeeping genes that are not under immune selection. As a result, sequence variation at these genetic loci is infrequent, explaining the low resolution obtained in our study for these MLST schemes. We also demonstrated that the MST (Klint *et al.*, 2007) system had the highest resolution when applied to the LGV strains in Chapter 4. This supports the findings of studies by Christerson *et al.* (2010), that show that this system is highly discriminatory. Our study identified that the MLVA-*ompA* genotyping system had the highest typeability of all four genotyping systems, and had a discriminatory capacity of 0.59, which was lower than that obtained in other MLVA-*ompA* genotyping studies (0.94 – 0.99) (Pedersen *et al.*, 2008; Ikryannikova *et al.*, 2010; Labiran *et al.*, 2017). A study by Labiran *et al.* (2016) in Brighton indicated that a high level of genotypic diversity within L2b isolates existed in Brighton. Hence, we determined that the LGV isolates included in the in silico study may not have been representative of the genotypic diversity that exists within LGV strains. We assessed the practicability of the MLVA-*ompA* genotyping method by utilising MLVA-*ompA* to genotype a rectal isolate obtained from an MSM in the Brighton (2016) study. This field-test also provided a platform to assess the reproducibility of the MLVA-*ompA* system, as Bri038 had previously been genotyped by Labiran *et al.* (2016). We obtained an identical MLVA-*ompA* genotype to that assigned to the isolate in 2016, and this demonstrated that the MLVA-*ompA* system is reproducible. Other studies also determined that the system is highly reproducible (Pedersen *et al.,* 2008; Ikryannikova *et al.*, 2010).

Over the past decade, a resurgence in LGV cases caused by *ompA* genotype L2 has been reported in Austria and Croatia (Cole *et al.*, 2020a), Spain (Rodríguez-Domínguez *et al.*, 2014), France (Peuchant *et al.*, 2016) and Sweden (Isaksson *et al.*, 2017); and in these countries, strains with *ompA* genotype L2 now dominate. In contrast, *ompA* genotyping of LGV samples from the UK (n= 22) (Cole *et al.*, 2020b) in 2020 showed that the L2b, and not L2, *ompA* genotype predominated. The Cole *et al.* (2020b) study included only a small number of LGV samples from the UK. No large-scale genotyping survey of LGV strains applying the high-resolution MLVA-*ompA* genotyping method had been conducted in the UK prior to our study in Chapter 5, and our understanding of the UK LGV epidemic previously relied on smaller studies utilising *ompA* genotyping (Cole *et al.*, 2020b). As

London has the highest incidence of LGV in the UK, it was decided to conduct a genotyping survey of LGV strains in London (**Chapter 5**). We obtained LGV-positive DNA extracts from high-risk groups (predominantly MSM) attending eight London GUM clinics, and utilised the MLVA-*ompA* genotyping system to assign genotypes to these DNA extracts. We demonstrated that the predominant *ompA* genotype within the London population was the L2 *ompA* genotype, and not the L2b *ompA* genotype that had been reported in UK populations since 2005 (Macdonald *et al.*, 2005). This was the first report of the predominance of L2 in the UK. We also found that two dominant VNTR types were co-circulating in London – 1.9.2b, and 1.9.3b – and these were associated with several distinct *ompA* genotypes. A previous MLVA-*ompA* genotyping study of rectal *C. trachomatis* strains in Brighton (2016) reported only 3 cases of 1.9.2b-L2b and no cases with 1.9.3b-L2b. These regional differences in genotypes could be the result of distinct dissemination patterns within each population, or the differences could represent a temporal shift in genotypes, given that our London study genotyped specimens collected in 2018 and 2019, and the Brighton study took place between 2011 and 2013. We detected a high level of diversity within *ompA* in the London study population. A high level of *ompA* diversity in LGV strains has previously been reported by Stary *et al.* (2008) and Cole *et al.* (2020b). Cole *et al.* (2020b) detected the novel L2b variants, L2bV5 and L2bV6, in four samples and one sample respectively from the UK. In contrast, our study reported a higher frequency of these variants, with L2bV5 and L2bV6 detected in 18 and 6 DNA extracts respectively. Our study has increased knowledge of the current LGV UK epidemic in MSM. We have also provided the first UK report of a hybrid L2b/D *ompA* profile previously detected in Portugal. Little is known about the pathogenic capability of this hybrid strain, and as a result, the public health implications of the L2b/D recombinant strain are as of yet unknown.

Whilst multi-locus genotyping systems allow genotypic diversity within *C. trachomatis* strains circulating within high-risk sexual networks to be elucidated; the resolution provided by these systems does not match that achieved by whole genome sequencing. Whole genome sequencing has provided insight into the genomic diversity of *C. trachomatis* strains (Jeffrey *et al.*, 2010; Harris *et al.*, 2012; Hadfield *et al.*, 2017). Whilst genomic analyses have been conducted on LGV strains isolated from patients with symptomatic LGV infections (i.e. the inguinal LGV syndrome or LGV proctitis) (Thomson *et al.*, 2007; Somboonna *et al.*, 2011; Borges and Gomes, 2015a); strains isolated from patients with asymptomatic LGV infections have not previously been analysed. It is unknown whether the absence of disease symptoms is the result of genetic variation within the genome, or a host effect. In Chapter 6, we applied whole genome sequencing to three LGV isolates obtained from symptomatic (n = 2) and asymptomatic (n=1) MSM. The study hypothesis was that sequence differences at genomic loci between isolates from symptomatic and asymptomatic MSM

in Brighton could account for the difference in patient symptomology. We showed that genomic diversity was present within the LGV isolates, and identified that all three isolates possessed an L2b genomic backbone. A key finding of the study was that two isolates (Bri088 and Bri098) contained a SNP that reverted the *ompA* genotype from L2b to L2, resulting in an L2b genomic backbone and an *ompA* genotype L2. This reversion SNP confirms the findings previously noted in Seth-Smith *et al.* (2021). In the Seth-Smith (2021) paper, 26 LGV genomes with *ompA* genotype L2 and an L2b genomic backbone were identified, all of which contained the SNP reversion detected in our study. In Chapter 6, we also identified that the two isolates obtained from symptomatic MSM (and not the isolate obtained from an asymptomatic MSM) contained SNPs within genes thought to play functional roles in intracellular replication and transition of EBs to RBs. The asymptomatic isolate (Bri098) contained two SNPs within hypothetical proteins, and these SNPs were not present in Bri038 or Bri088. These SNPs were all non-synonymous. It was unclear whether mutations within the identified genes contributed to the difference in symptomology exhibited by the three patients, and this was identified as a limitation of the study.

This PhD project has led to multiple avenues for further research. The key question that remains to be addressed is whether the genomic variation identified within the three LGV isolates in Chapter 6 contributed to the symptomatic or asymptomatic phenotype in the host. Potential next steps for this work could include obtaining a larger sample size of LGV isolates from patients with a symptomatic versus asymptomatic clinical manifestation. This would provide more assurance that variation within isolates is representative of the population. Comparative genomic analysis of each isolate could be combined with in vitro experimentation in order to test whether the genetic variation results in different pathogenic properties within isolates, and whether these can be correlated with clinical manifestation. A possible avenue for research could be to observe growth rate and inclusion morphology by visualising under phase contrast at various time points after infection.

APPENDIX 1 Publication One: *OmpA* Genotyping of a Conjunctival Swab Obtained from a Child

Case report

Reducing suspicion of sexual abuse in paediatric chlamydial conjunctivitis using *ompA* genotyping

Alexander Mitchell,[1] Mitul Patel,[2] Chloe Manning,[3] Joseph Abbott[1]

[1]Ophthalmology, Birmingham Children's Hospital, Birmingham, UK
[2]Microbiology, Birmingham Children's Hospital, Birmingham, UK
[3]Faculty of Medicine, University of Southampton, Southampton, UK

Correspondence to
Dr Alexander Mitchell;
alexander.mitchell1@nhs.net

Accepted 19 March 2021

SUMMARY

Chlamydia trachomatis is a Gram-negative bacterium that causes urogenital tract infections, and ocular infections including trachoma, neonatal conjunctivitis and adult chlamydial inclusion conjunctivitis. A positive *C. trachomatis* diagnosis in children often raises suspicions of sexual abuse. While outer membrane protein A (*ompA*) genotypes A–C are non-invasive and are associated with trachoma; *ompA* genotypes D–K are often associated with sexually transmitted urogenital infections or sexually acquired chlamydial conjunctivitis. A 10-year-old female presented with a 7-month history of unilateral conjunctivitis with itching, watering and hyperaemia. She had recently moved from an urban centre in Afghanistan to the UK. A conjunctival swab taken from the child tested positive for *C. trachomatis*. Application of *ompA* genotyping to conjunctival swab chlamydial DNA demonstrated that the *C. trachomatis* had an *ompA* genotype C. Chlamydial strains with this *ompA* genotype cause trachoma and have never previously been associated with urogenital infection. This result supported cessation of child protection investigations.

BACKGROUND

Chlamydia trachomatis is an obligate intracellular Gram-negative bacterium with a biphasic developmental cycle that alternates between infectious extracellular elementary bodies, and non-infectious intracellular reticulate bodies.[1] It primarily infects ocular and genital tract epithelial cells. Chlamydial infection of ocular epithelial cells can result in trachoma, neonatal conjunctivitis and adult chlamydial inclusion conjunctivitis.[2]

The outer membrane protein A (*ompA*) gene encodes the major outer membrane protein of *C. trachomatis*. This gene exhibits extensive nucleotide sequence variation that is localised to four discrete regions, termed variable segments I to IV (VS I–IV). *ompA* genotyping allows for chlamydial strain differentiation.[3] Strains with *ompA* genotypes A–C cause the blinding disease trachoma, which is the leading infectious cause of blindness worldwide. Trachoma is a public health problem in 44 countries, particularly in rural regions of Africa, Central and South America, Asia, Australia and the Middle East.[4] *C. trachomatis* strains with *ompA* genotypes D–K can cause urogenital infections and adult inclusion conjunctivitis. Chlamydial inclusion conjunctivitis is a chronic mucopurulent follicular conjunctivitis which is often unilateral, with associated lid oedema and non-tender lymphadenopathy. The cornea can show punctate epithelial erosions, subepithelial opacities and marginal infiltrates. It can be challenging to distinguish from other forms of conjunctivitis initially and is often diagnosed from conjunctival swabs when clinical suspicion has been raised.

CASE PRESENTATION

A 10-year-old female presented to the Ophthalmology Department via her optician describing a 7-month history of persistent left itchy, watery eye and mild lid swelling. She had travelled by land to the UK from an urban centre in Afghanistan 8 months previously. She had no significant medical history or previous eye complaints and had no known allergies. A recent course of chloramphenicol eye drops had been found to make no subjective difference to symptoms or appearance of the eye.

The visual acuity in the right eye was 0.360, left eye 0.380 logMAR visual acuity unaided. The left conjunctiva showed superficial punctate stain and superior injection with fine corneal vascularisation and pannus. There was hyperaemia and papillae of the lower lid subtarsal conjunctiva, with raised fleshy lesions in the inferior fornix, with mild follicular reaction of the superior tarsal conjunctiva. Subepithelial infiltrates were found throughout the central and inferior cornea with no associated epitheliopathy on fluorescein staining. The anterior chamber was deep with no inflammatory cells seen, and the rest of her ocular examination was within normal limits, with a healthy, unaffected right eye. Cycloplegic retinoscopy showed an uncorrected small hypermetropia with astigmatism, and spectacles were prescribed. Findings were atypical for conjunctivitis and given the social context, a conjunctival swab of the left eye was taken and tested for *C. trachomatis* and *Neisseria gonorrhoeae*.

INVESTIGATIONS

The swabs taken from the left eye were tested for *C. trachomatis* and *N. gonorrhoeae* using the real-time CT/NG combined PCR assay (Abbott). The swab tested positive for *C. trachomatis* and negative for *N. gonorrhoeae*. Formal child protection investigations were undertaken by a paediatrician with expertise in child abuse but did not include genital examination or screening for other sexually transmitted infections. The patient denied any inappropriate intimate contact, with no genital pain, vaginal bleeding or discharge. The parents had not knowingly been infected with Chlamydia and denied having any symptoms.

A left conjunctival swab was sent to the Department of Molecular Microbiology at Southampton

To cite: Mitchell A, Patel M, Manning C, *et al*. *BMJ Case Rep* 2021;**14**:e238871. doi:10.1136/bcr-2020-238871

Figure 1 Slit-lamp photograph showing resolved corneal vascularisation and fine subepithelial corneal infiltrates centrally and inferiorly following treatment.

General Hospital for *ompA* genotyping. Chlamydial genomic DNA was isolated from the swab using NucleoSpin DNA Trace Tissue Kit (Macherey-Nagel, Germany). *ompA* was amplified using primers PCTM3 and NR1.[5] PCR was carried out in a 20 µL volume consisting of 10 µL Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, UK), 500 nM of the forward and reverse primers (Eurogentec, Belgium) and 1 µL of DNA. PCR amplification was conducted using a Veriti Thermal Cycler (Applied Biosystems, UK). The PCR conditions were as follows: initial denaturation at 98°C for 10 s, 40 cycles of 98°C for 2 s, 59°C for 5 s and 72°C for 10 s; and a final extension at 72°C for 1 min. PCR product was purified using the Wizard SV Gel and PCR Clean-Up System (Promega, UK), according to manufacturer's instructions. PCR amplicons (1 ng/µL/100 bp) were commercially sequenced at Source Bioscience (Cambridge, UK). An alphabetical *ompA* genotype was assigned to the sample via *ompA* sequence comparison to the NCBI database using BLAST.

The sample shared the closest *ompA* sequence identity to strains with *ompA* genotype C, a known trachoma-causing *ompA* genotype that has never been associated with urogenital infection. This ruled out the possibility of acquisition of *C. trachomatis* via sexual routes and established the diagnosis of trachoma.

OUTCOME AND FOLLOW-UP

Following a 2-week course of oral erythromycin, corneal vascularisation and ocular hyperaemia regressed (figure 1). Symptoms improved subjectively and have been controlled to date with occasional preservative free lubricant eyedrops. Best subsequent visual acuities are 0.18 right eye, 0.5 left eye with spectacle correction 2 years following initial presentation. A coincidental and uncorrected astigmatism has caused the minor reduction in final best vision. Conjunctival swabs were taken from each eye following treatment, and each tested negative for *C. trachomatis*.

The child protection team were satisfied that this was not a case of sexual abuse, either intrafamilial or non-familial, that no further investigations were required, and child protection proceedings were halted.

DISCUSSION

Acute bacterial conjunctivitis is a common presentation to primary care and emergency eye services. Antibiotic treatments are often prescribed and have been shown to reduce the duration of symptoms and may have a role in limiting infection with minimal side effects.[6]

Perinatal transmission from the mother during delivery can cause neonatal conjunctival and respiratory tract infections.

Outside of the neonatal period, chlamydial conjunctivitis is classically described as a unilateral, persistent follicular conjunctivitis, often with mucopurulent discharge.[7 8] Between 0.2% and 3% of genital chlamydial infections are complicated by conjunctivitis and a large proportion of patients (up to 61% of men and up to 90% of women) presenting with chlamydial conjunctivitis have concomitant urogenital infection. Genital infections have often started to heal by the time of ocular involvement, thus in adults, rates of primary genital infection may be even higher than described.[9]

ompA genotyping is a useful tool for *C. trachomatis* strain discrimination. This method is relatively low cost and is performed in a research laboratory setting. In a survey of *C. trachomatis ompA* genotypes circulating in trachoma-endemic regions of Australia, 0 of the 217 genotyped urogenital specimens yielded trachoma *ompA* genotypes.[10]

In non-endemic regions, clinical suspicion of chlamydial infection is often raised in atypical presentations or when symptoms fail to respond to initial broad spectrum treatment. In prepubertal children investigated for child abuse, genital chlamydial infections were found in 3.1%.[11] Chlamydial conjunctivitis is uncommon outside of the neonatal period and may be the only presenting feature of child abuse. These patients are considered high risk and are referred for child protection investigation.

Current UK NICE guidelines recommend Nucleic Acid Amplification Test (NAAT) for urogenital and extragenital samples from adults,[12] and the British Association for Sexual Health and HIV recommends the use of NAAT to test for *C. trachomatis* in prepubertal children, recognising that availability of validated culture testing is extremely limited if available at all.[11] Concerns for the use of NAAT for *C. trachomatis* in prepubertal children include low positive predictive value in a low prevalence population. In the USA, NAAT is recommended when testing for *C. trachomatis* in all adults, however culture remains the preferred method of testing extragenital sites in boys and prepubertal girls in cases of suspected child abuse, due to insufficient data.[14]

While inappropriate inaction can lead to tragic failures, unwarranted investigations can lead to unnecessary, costly and disproportionate interventions which can damage trust and communication between families and the professionals they interact with.

Learning points

- Chlamydial conjunctivitis beyond the neonatal period can be the only presenting feature of child abuse.
- Outer membrane protein A (*ompA*) gene sequence analysis can be applied to differentiate between trachoma-causing strains and sexually transmitted urogenital strains of *Chlamydia trachomatis*.
- Exclusion of *ompA* genotypes associated with urogenital infection has profound implications for child protection investigations.

As the infection originated in Afghanistan, where collection of epidemiological data on trachoma has been limited by conflict, this may represent one of the first reported confirmed trachoma-causing strains from that country to be *ompA* genotyped.

While the lack of genital chlamydia swab investigation, as deemed not necessary by the child protection team, may represent a limitation of our conclusions; the trachoma-causing *ompA* genotypes of *C. trachomatis* have not previously been associated with child sexual abuse. In this case, *ompA* genotype determination had a profound impact and supported the suspension of investigations by the child protection team

Acknowledgements The authors thank Professor Ian Clarke, Department of Molecular Microbiology, University of Southampton. Dr Helen Morris, Consultant Paediatrician, Birmingham Women's and Children's Hospitals.

Contributors Report was written by AM with support of JA and MP. Patient was under the care of JA. CM provided *ompA* gene sequence analysis and manuscript editing. Microbiology information provided by MP.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Parental/guardian consent obtained.

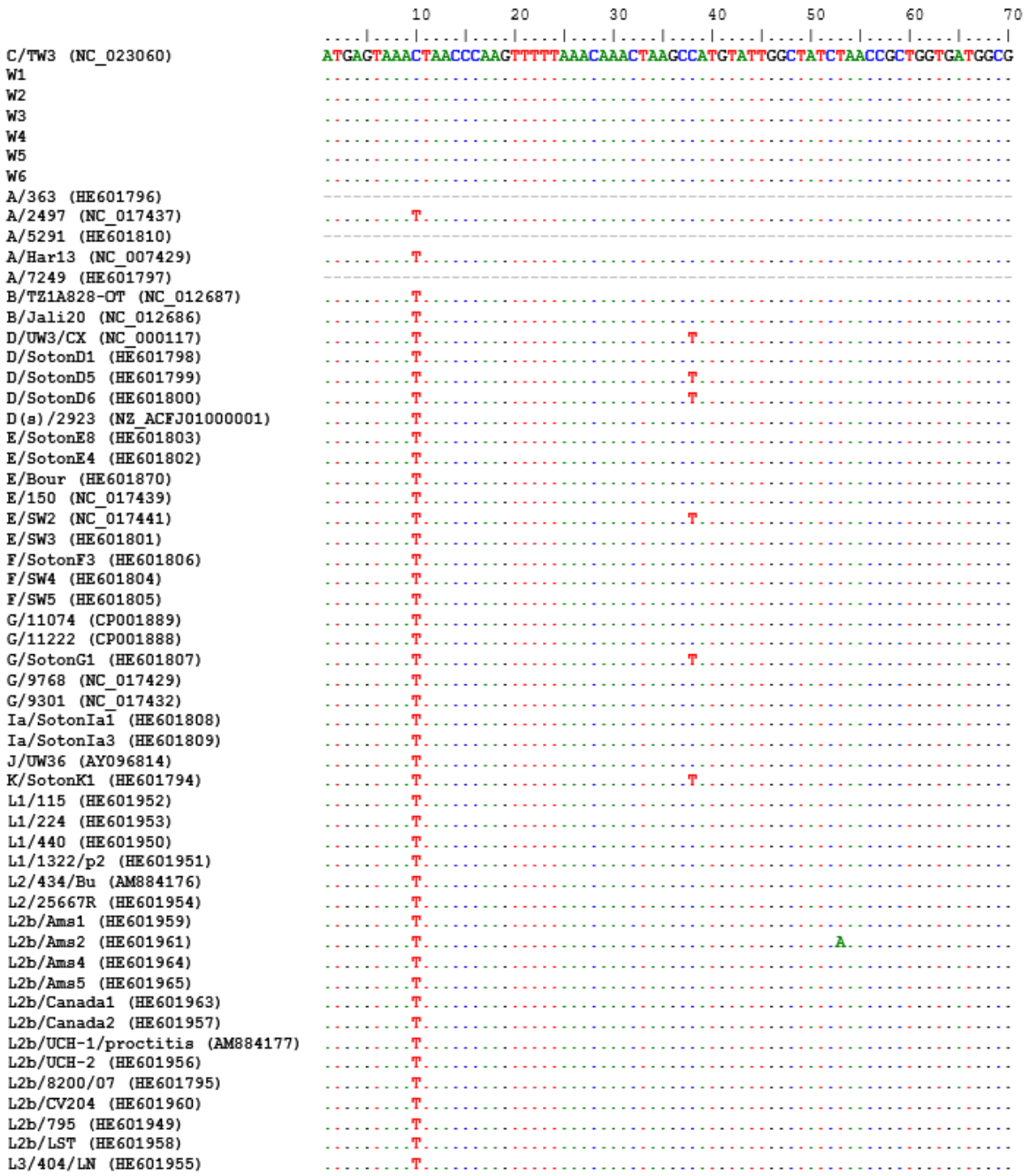
Provenance and peer review Not commissioned; externally peer reviewed.

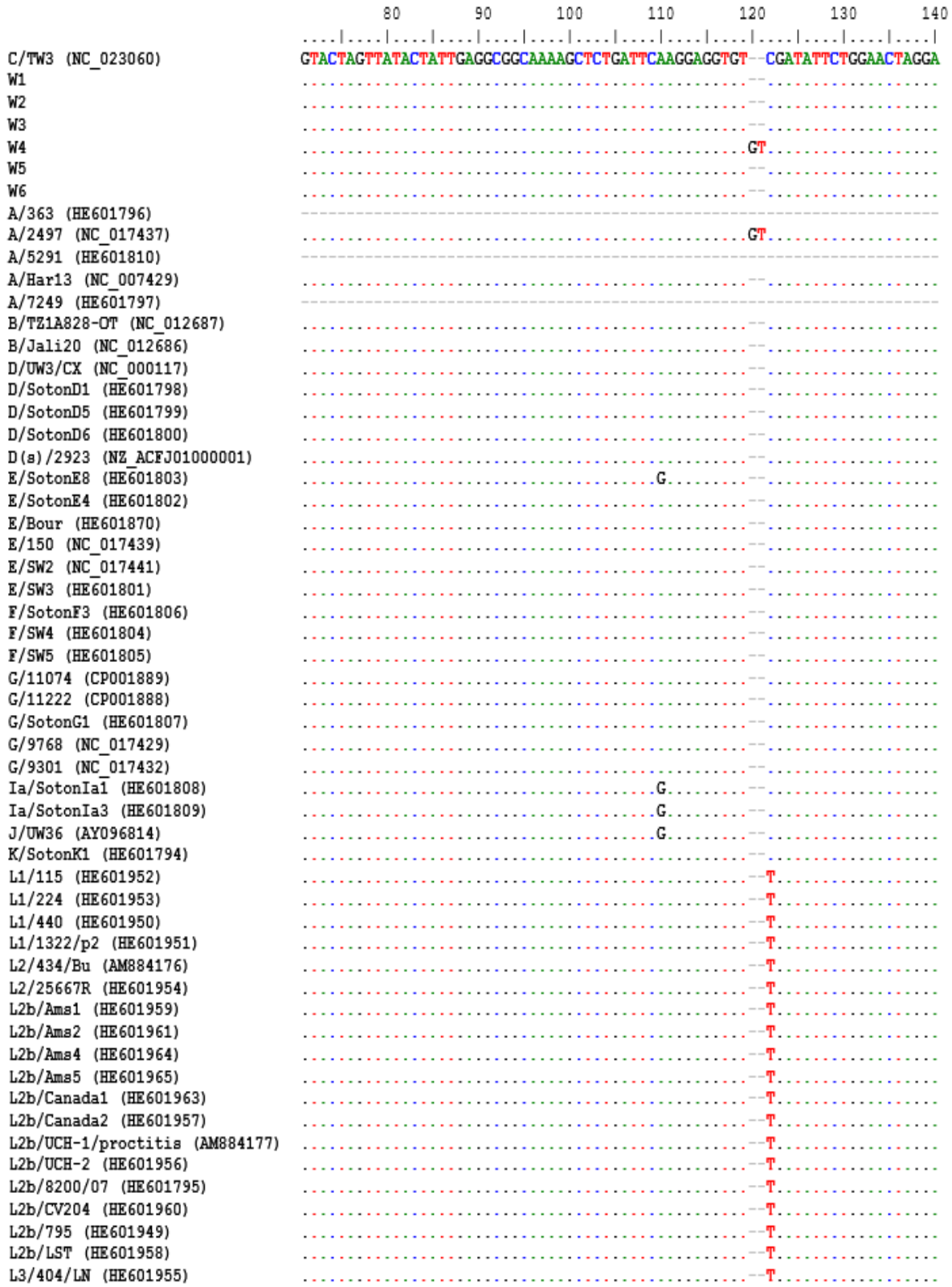
REFERENCES

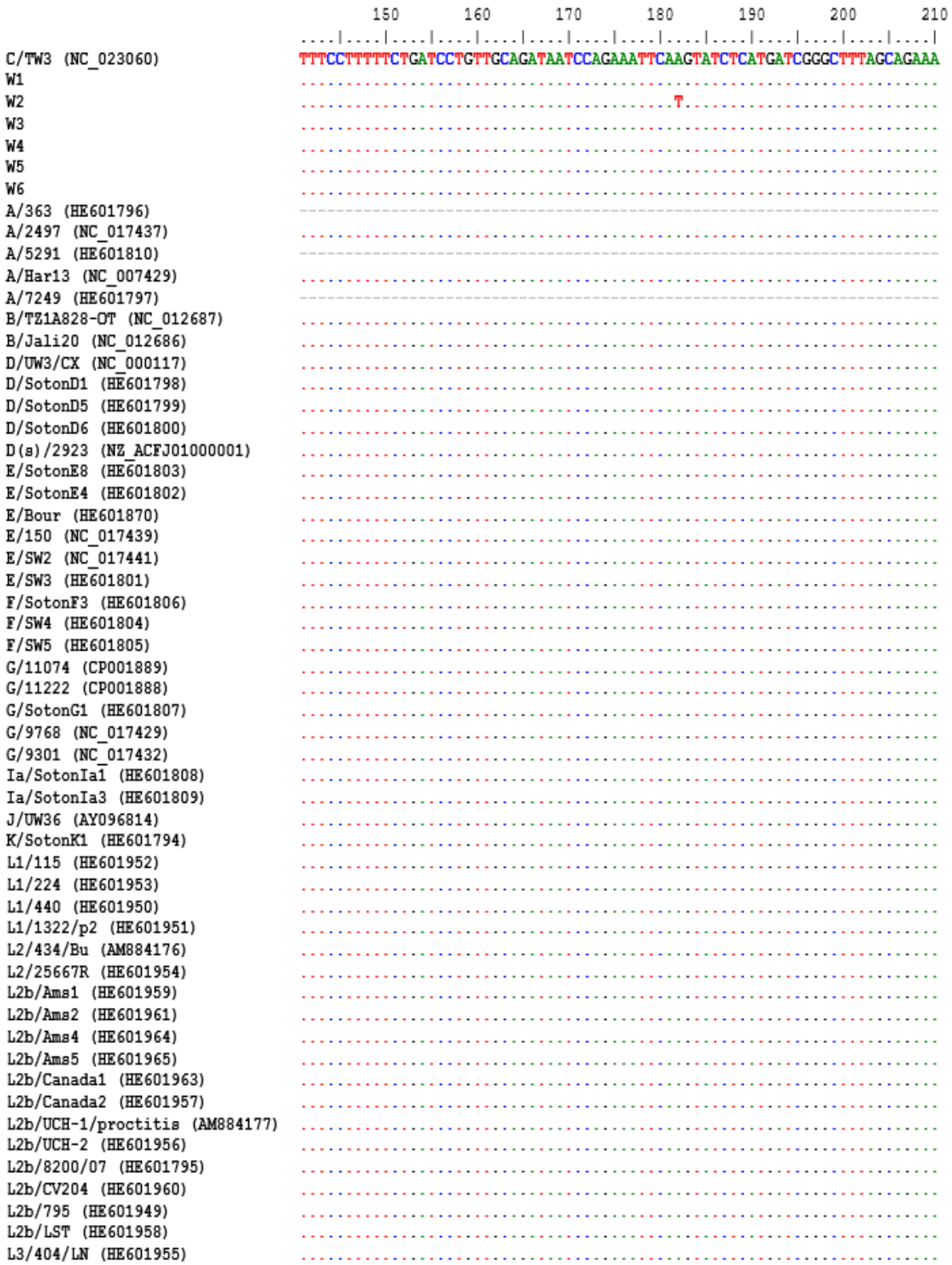
1 Elwell C, Mirrashidi K, Engel J. Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol* 2016;14:385–400.
2 Satpathy G, Behera HS, Ahmed NH. Chlamydial eye infections: current perspectives. *Indian J Ophthalmol* 2017;65:97–102.
3 Rawre J, Juyal D, Dhawan B. Molecular typing of *Chlamydia trachomatis*: an overview. *Indian J Med Microbiol* 2017;35:17–26.
4 WHO. World Health Organisation fact sheet: trachoma, 2020. Available: https://www.who.int/news-room/fact-sheets/detail/trachoma; [Accessed 20 May 2020].
5 Lan J, Ossewaarde JM, Walboomers JM, *et al*. Improved PCR sensitivity for direct genotyping of Chlamydia trachomatis serovars by using a nested PCR. *J Clin Microbiol* 1994;32:528–30.
6 Sheikh A, Hurwitz B, van Schayck CP, *et al*. Antibiotics versus placebo for acute bacterial conjunctivitis. *Cochrane Database Syst Rev* 2012;145:CD001211.
7 Leibowitz HM. The red eye. *N Engl J Med* 2000;343:345–51.
8 Brownell AD, Miraldi Utz V, Makoroff KL, *et al*. Chlamydia conjunctivitis in 2 prepubertal children: an uncommon presentation of child sexual abuse. *Pediatr Emerg Care* 2020;36:e473–5.
9 Postema EJ, Remeijer L, van der Meijden WI. Epidemiology of genital chlamydial infections in patients with chlamydial conjunctivitis; a retrospective study. *Sex Transm Infect* 1996;72:203–5.
10 Gilfard PM, Brenner NC, Tabrizi SN, *et al*. *Chlamydia trachomatis* genotypes in a cross-sectional study of urogenital samples from remote Northern and Central Australia. *BMJ Open* 2016;6:e009624.
11 Girardet RG, Lahoti S, Howard LA, *et al*. Epidemiology of sexually transmitted infections in suspected child victims of sexual assault. *Pediatrics* 2009;124:79–86.
12 Nwokolo NC, Dragovic B, Patel S, *et al*. 2015 UK national guideline for the management of infection with Chlamydia trachomatis. *Int J STD AIDS* 2016;27:251–67.
13 Rogstad K, Thomas A, Williams O. United Kingdom national guideline on the management of sexually transmitted infections and related conditions in children and young people – 2010; clinical effectiveness group, British association for sexual health and HIV. Available: https://www.bashhguidelines.org/media/1081/2674.pdf [Accessed 14 Jan 2021].
14 CDC. STD treatment guidelines. MMWR, Vol. 64(RR3), 2015. Available: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr6403a1.htm [Accessed 14 Jan 2021].

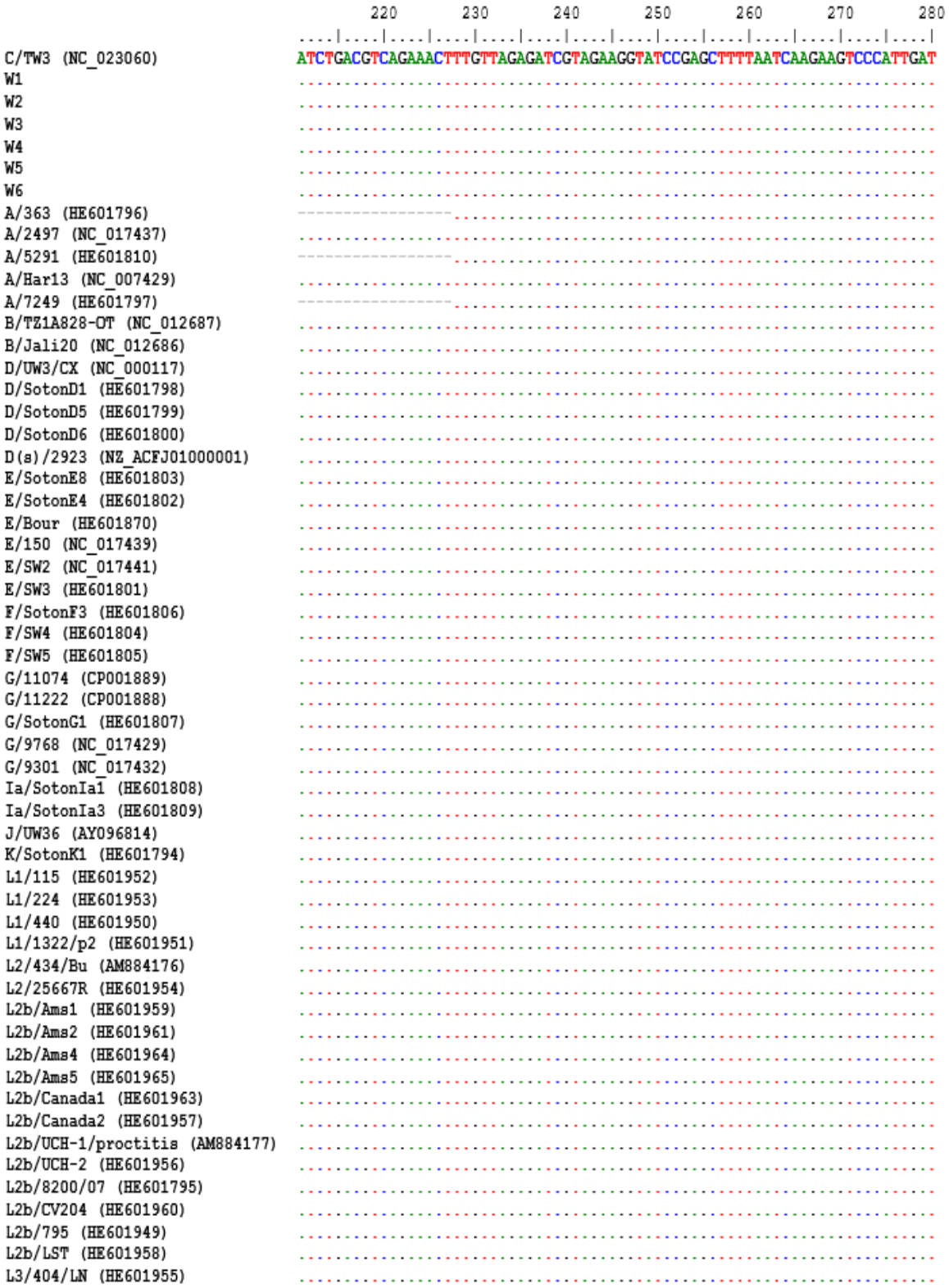
APPENDIX 2 Alignment of the *C. trachomatis trpA* gene from six white colonies (W1-W6)

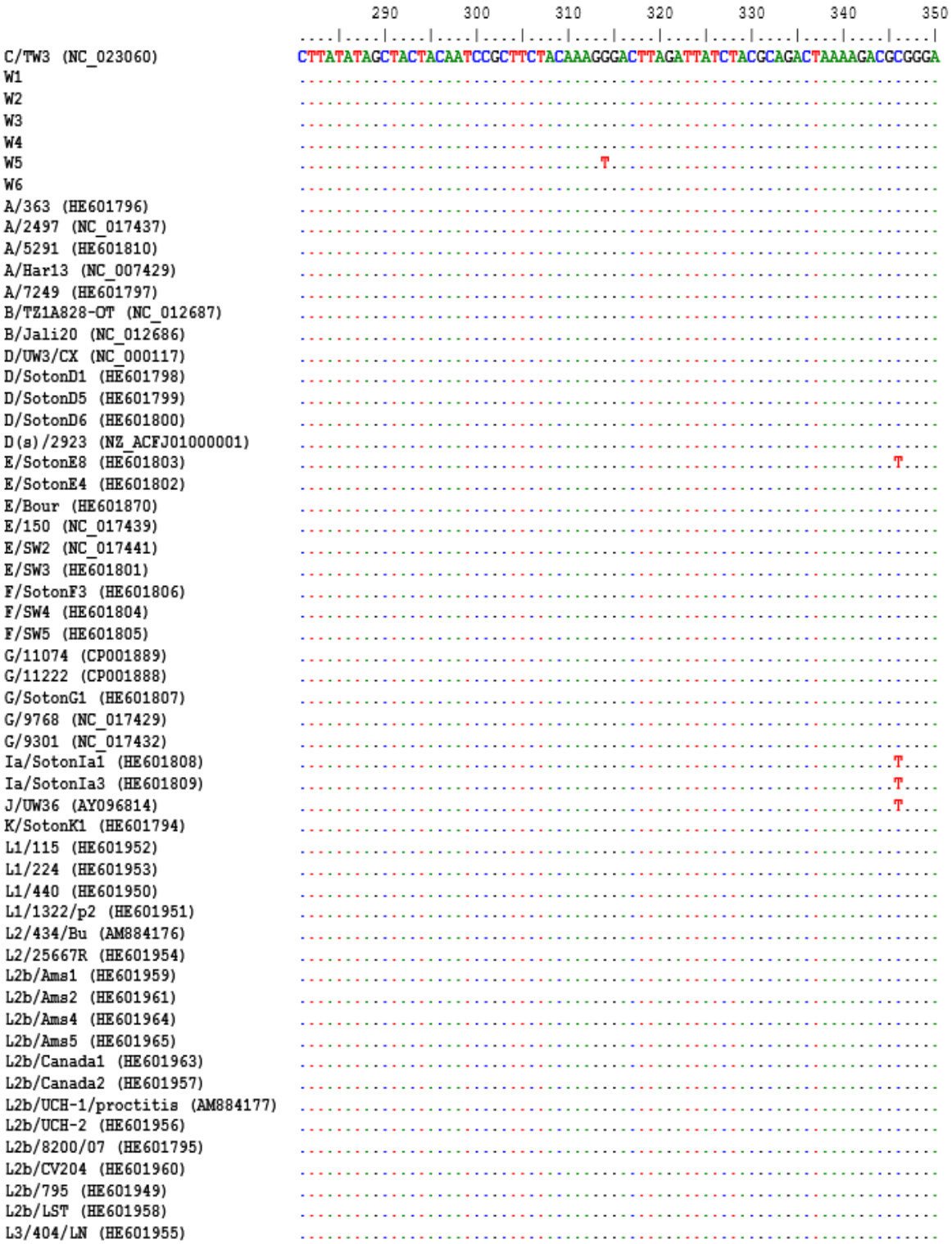
Sequence alignment of tryptophan synthase alpha (*trpA*) gene from six white colonies, W1 - W6. The *trpA* sequence of C/TW3 (Genbank Accession No. NC_023060.1) was used as a reference. Nucleotide numbers are given according to C/TW3. *TrpA* sequences extracted from fifty *C. trachomatis* strains were also included in the alignment, Genbank Accession Numbers are provided in brackets for each strain. Coloured dots indicate bases that were identical to C/TW3.

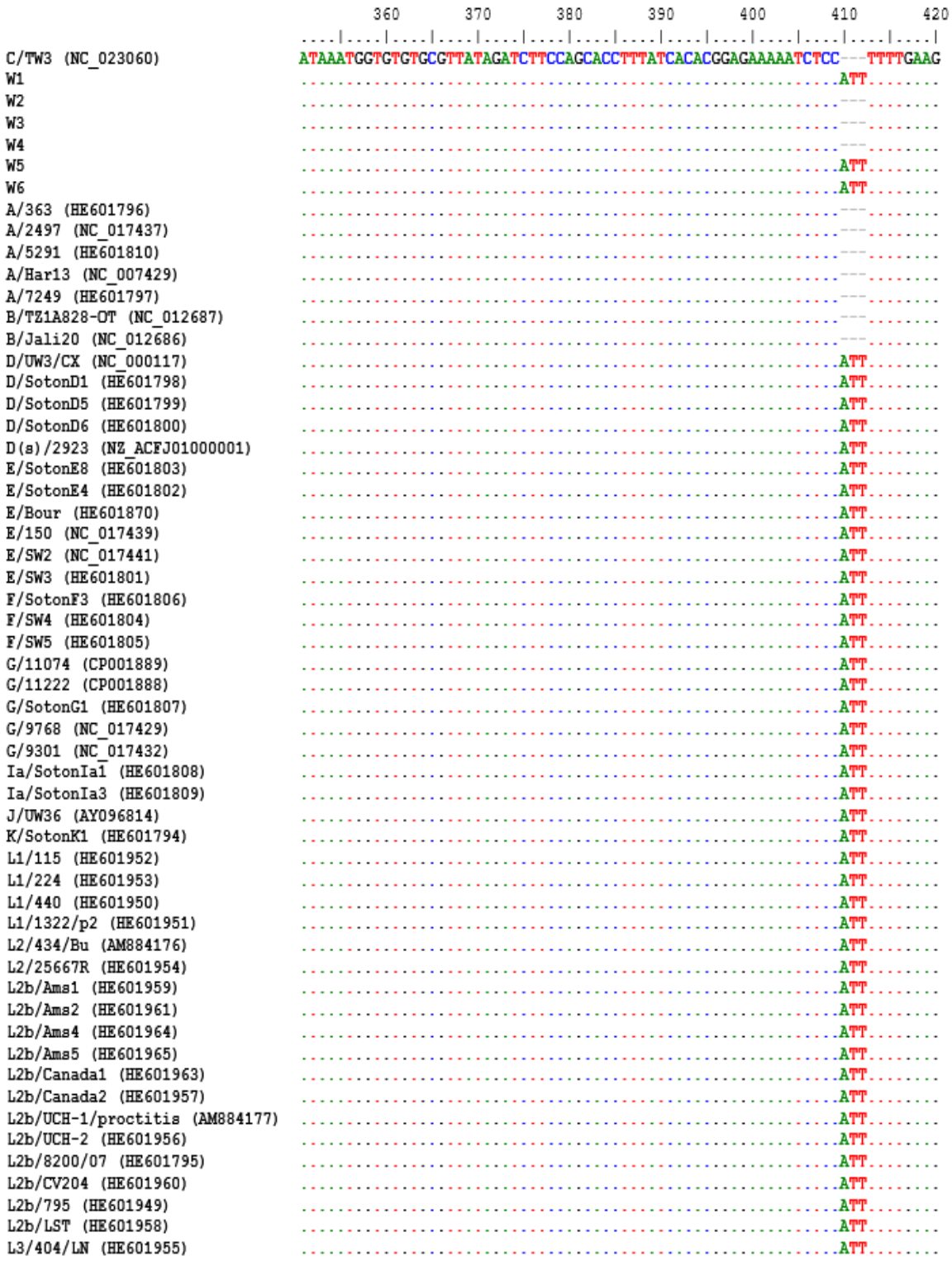


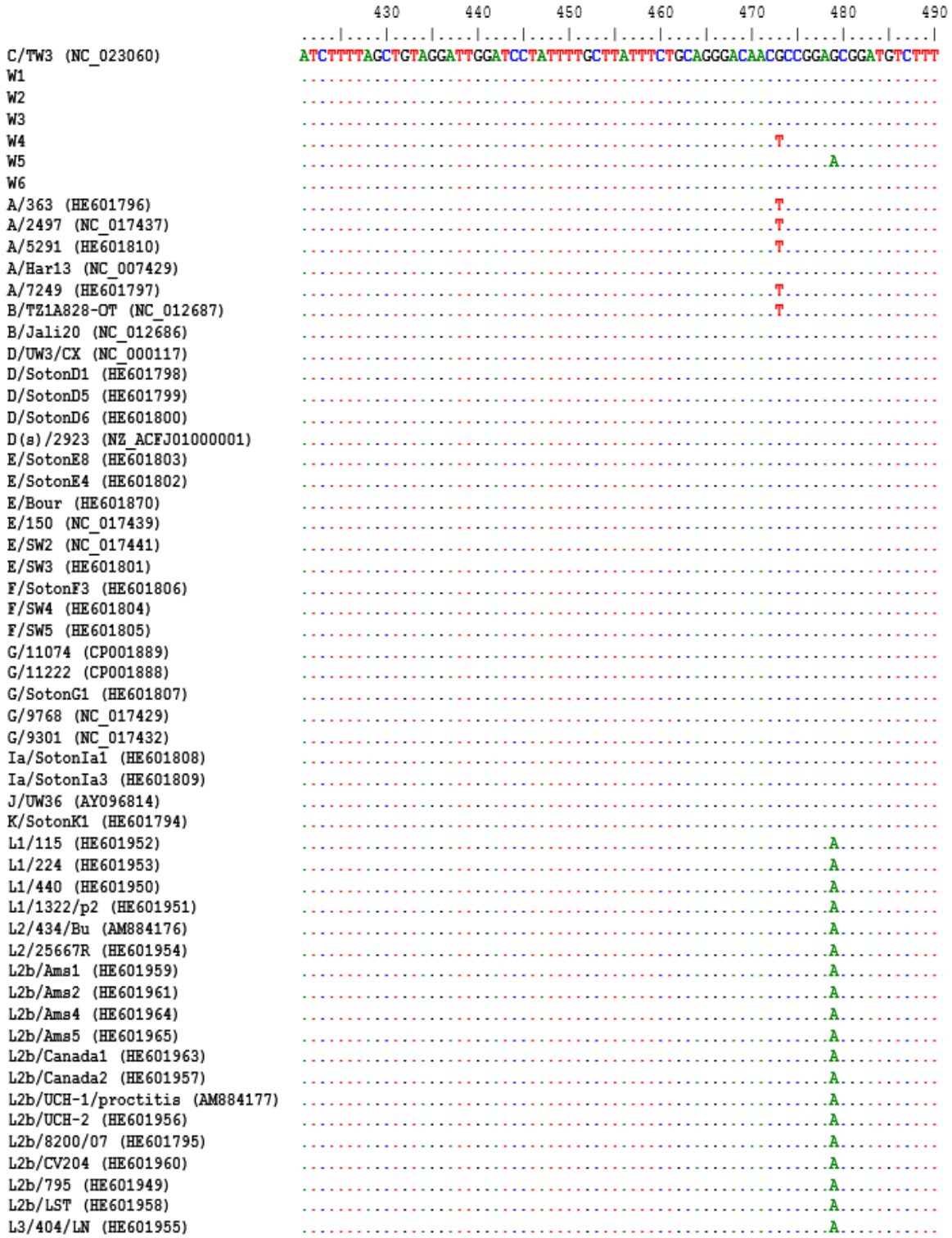

80 90 100 110 120 130 140
C/TW3 (NC_023060) GTACTAGTTATACTATTGAGGCGGCAAAAGCTCTGATTCAAGGAGGTGT--CGATATTCTGGAACTAGGA
W1
W2
W3
W4 GT
W5
W6
A/363 (HE601796)
A/2497 (NC_017437) GT
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803) G
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808) G
Ia/SotonIa3 (HE601809) G
J/UW36 (AY096814) G
K/SotonK1 (HE601794)
L1/115 (HE601952) T
L1/224 (HE601953) T
L1/440 (HE601950) T
L1/1322/p2 (HE601951) T
L2/434/Bu (AM884176) T
L2/25667R (HE601954) T
L2b/Ams1 (HE601959) T
L2b/Ams2 (HE601961) T
L2b/Ams4 (HE601964) T
L2b/Ams5 (HE601965) T
L2b/Canada1 (HE601963) T
L2b/Canada2 (HE601957) T
L2b/UCH-1/proctitis (AM884177) T
L2b/UCH-2 (HE601956) T
L2b/8200/07 (HE601795) T
L2b/CV204 (HE601960) T
L2b/795 (HE601949) T
L2b/LST (HE601958) T
L3/404/LN (HE601955) T

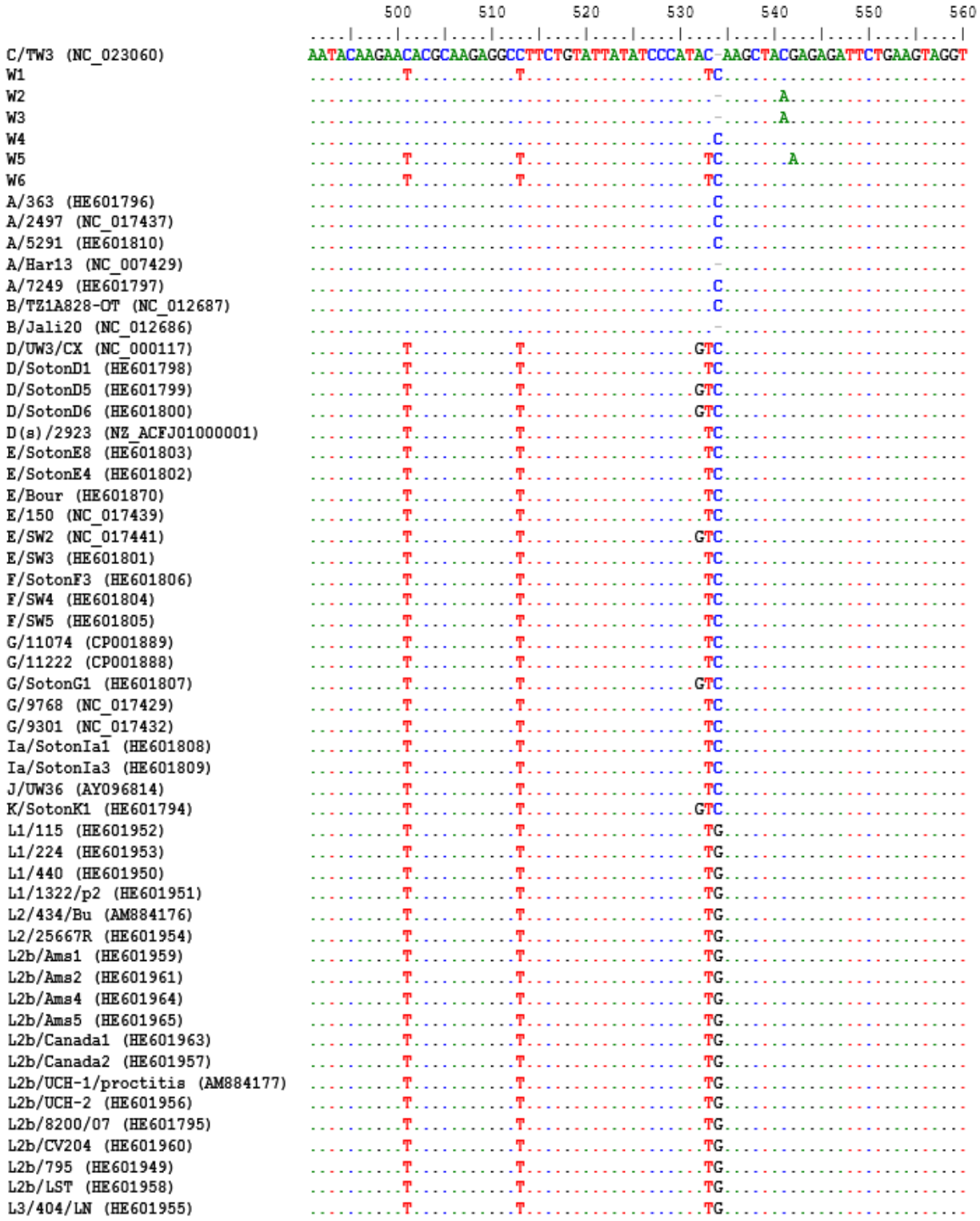

150 160 170 180 190 200 210
C/TW3 (NC_023060) TTTCCTTTTTCTGATCCTGTTGCAGATAATCCAGAAATTCAAGTATCTCATGATCGGGCTTTAGCAGAAA
W1
W2 T
W3
W4
W5
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808)
Ia/SotonIa3 (HE601809)
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)

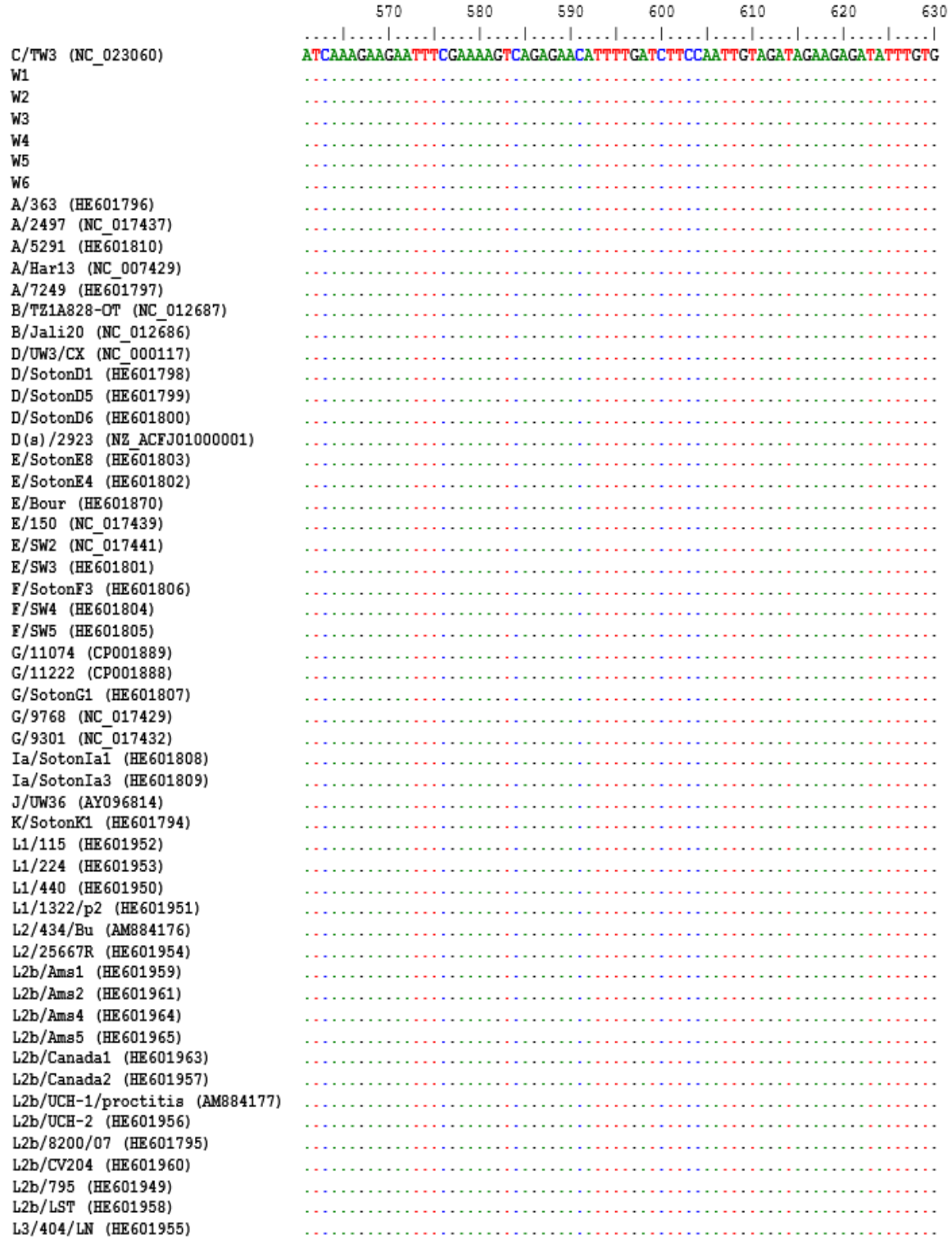

220 230 240 250 260 270 280
C/TW3 (NC_023060) ATCTGACGTCAGAAACTTTGTTAGAGATCGTAGAAGGTATCCGAGCTTTTAATCAAGAAGTCCCATTGAT
W1
W2
W3
W4
W5
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808)
Ia/SotonIa3 (HE601809)
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)

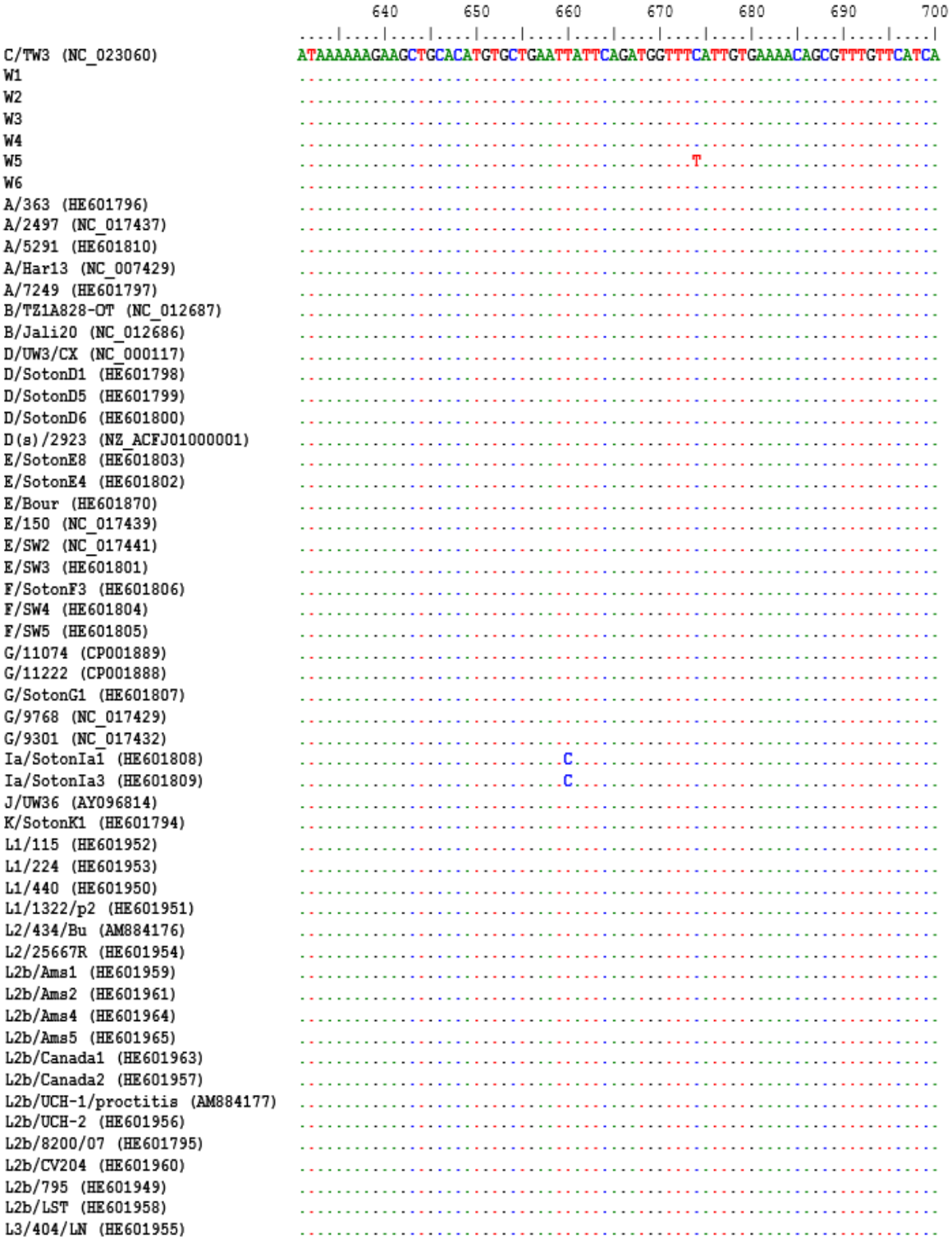

290 300 310 320 330 340 350
C/TW3 (NC_023060) CTTATATAGCTACTACAATCCGCTTCTACAAAGGGACTTAGATTATCTACGCAGACTAAAAGACGCGGGA
W1
W2
W3
W4
W5 T
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803) T
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808) T
Ia/SotonIa3 (HE601809) T
J/UW36 (AY096814) T
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)

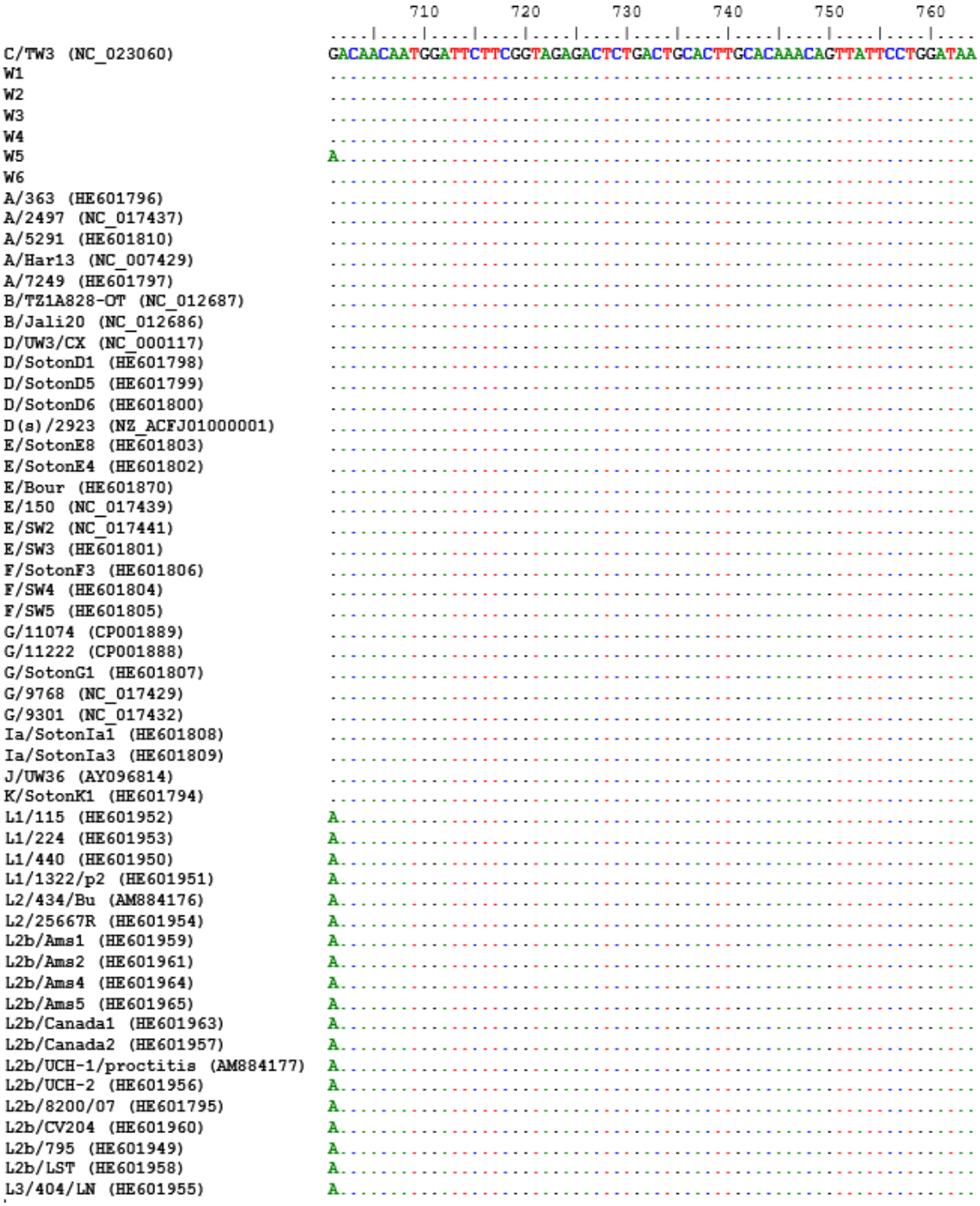

360 370 380 390 400 410 420
C/TW3 (NC_023060) ATAAATGGTGTGTGCGTTATAGATCTTCCAGCACCTTTATCACACGGAGAAAAATCTCC---TTTTGAAG
W1 ATT
W2 ---
W3 ---
W4 ---
W5 ATT
W6 ATT
A/363 (HE601796) ---
A/2497 (NC_017437) ---
A/5291 (HE601810) ---
A/Har13 (NC_007429) ---
A/7249 (HE601797) ---
B/TZ1A828-OT (NC_012687) ---
B/Jali20 (NC_012686) ---
D/UW3/CX (NC_000117) ATT
D/SotonD1 (HE601798) ATT
D/SotonD5 (HE601799) ATT
D/SotonD6 (HE601800) ATT
D(s)/2923 (NZ_ACFJ01000001) ATT
E/SotonE8 (HE601803) ATT
E/SotonE4 (HE601802) ATT
E/Bour (HE601870) ATT
E/150 (NC_017439) ATT
E/SW2 (NC_017441) ATT
E/SW3 (HE601801) ATT
F/SotonF3 (HE601806) ATT
F/SW4 (HE601804) ATT
F/SW5 (HE601805) ATT
G/11074 (CP001889) ATT
G/11222 (CP001888) ATT
G/SotonG1 (HE601807) ATT
G/9768 (NC_017429) ATT
G/9301 (NC_017432) ATT
Ia/SotonIa1 (HE601808) ATT
Ia/SotonIa3 (HE601809) ATT
J/UW36 (AY096814) ATT
K/SotonK1 (HE601794) ATT
L1/115 (HE601952) ATT
L1/224 (HE601953) ATT
L1/440 (HE601950) ATT
L1/1322/p2 (HE601951) ATT
L2/434/Bu (AM884176) ATT
L2/25667R (HE601954) ATT
L2b/Ams1 (HE601959) ATT
L2b/Ams2 (HE601961) ATT
L2b/Ams4 (HE601964) ATT
L2b/Ams5 (HE601965) ATT
L2b/Canada1 (HE601963) ATT
L2b/Canada2 (HE601957) ATT
L2b/UCH-1/proctitis (AM884177) ATT
L2b/UCH-2 (HE601956) ATT
L2b/8200/07 (HE601795) ATT
L2b/CV204 (HE601960) ATT
L2b/795 (HE601949) ATT
L2b/LST (HE601958) ATT
L3/404/LN (HE601955) ATT


430 440 450 460 470 480 490
C/TW3 (NC_023060) ATCTTTTAGCTGTAGGATTGGATCCTATTTTGCTTATTTCTGCAGGGACAACGCCGGAGCGGATGTCTTT
W1
W2
W3
W4
W5
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808)
Ia/SotonIa3 (HE601809)
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)


500 510 520 530 540 550 560
C/TW3 (NC_023060) AATACAAGAACACGCAAGAGGCCTTCTGTATTATATCCCATAC-AAGCTACGAGAGATTCTGAAGTAGGT
W1 .....T.....T.....TC.....
W2 .....-.....A.....
W3 .....-.....A.....
W4 .....C.....
W5 .....T.....T.....TC.....A.....
W6 .....T.....T.....TC.....
A/363 (HE601796) .....C.....
A/2497 (NC_017437) .....C.....
A/5291 (HE601810) .....C.....
A/Har13 (NC_007429) .....-.....
A/7249 (HE601797) .....C.....
B/TZ1A828-OT (NC_012687) .....C.....
B/Jali20 (NC_012686) .....-.....
D/UW3/CX (NC_000117) .....T.....T.....GTC.....
D/SotonD1 (HE601798) .....T.....T.....TC.....
D/SotonD5 (HE601799) .....T.....T.....GTC.....
D/SotonD6 (HE601800) .....T.....T.....GTC.....
D(s)/2923 (NZ_ACFJ01000001) .....T.....T.....TC.....
E/SotonE8 (HE601803) .....T.....T.....TC.....
E/SotonE4 (HE601802) .....T.....T.....TC.....
E/Bour (HE601870) .....T.....T.....TC.....
E/150 (NC_017439) .....T.....T.....TC.....
E/SW2 (NC_017441) .....T.....T.....GTC.....
E/SW3 (HE601801) .....T.....T.....TC.....
F/SotonF3 (HE601806) .....T.....T.....TC.....
F/SW4 (HE601804) .....T.....T.....TC.....
F/SW5 (HE601805) .....T.....T.....TC.....
G/11074 (CP001889) .....T.....T.....TC.....
G/11222 (CP001888) .....T.....T.....TC.....
G/SotonG1 (HE601807) .....T.....T.....GTC.....
G/9768 (NC_017429) .....T.....T.....TC.....
G/9301 (NC_017432) .....T.....T.....TC.....
Ia/SotonIa1 (HE601808) .....T.....T.....TC.....
Ia/SotonIa3 (HE601809) .....T.....T.....TC.....
J/UW36 (AY096814) .....T.....T.....TC.....
K/SotonK1 (HE601794) .....T.....T.....GTC.....
L1/115 (HE601952) .....T.....T.....TG.....
L1/224 (HE601953) .....T.....T.....TG.....
L1/440 (HE601950) .....T.....T.....TG.....
L1/1322/p2 (HE601951) .....T.....T.....TG.....
L2/434/Bu (AM884176) .....T.....T.....TG.....
L2/25667R (HE601954) .....T.....T.....TG.....
L2b/Ams1 (HE601959) .....T.....T.....TG.....
L2b/Ams2 (HE601961) .....T.....T.....TG.....
L2b/Ams4 (HE601964) .....T.....T.....TG.....
L2b/Ams5 (HE601965) .....T.....T.....TG.....
L2b/Canada1 (HE601963) .....T.....T.....TG.....
L2b/Canada2 (HE601957) .....T.....T.....TG.....
L2b/UCH-1/proctitis (AM884177) .....T.....T.....TG.....
L2b/UCH-2 (HE601956) .....T.....T.....TG.....
L2b/8200/07 (HE601795) .....T.....T.....TG.....
L2b/CV204 (HE601960) .....T.....T.....TG.....
L2b/795 (HE601949) .....T.....T.....TG.....
L2b/LST (HE601958) .....T.....T.....TG.....
L3/404/LN (HE601955) .....T.....T.....TG.....

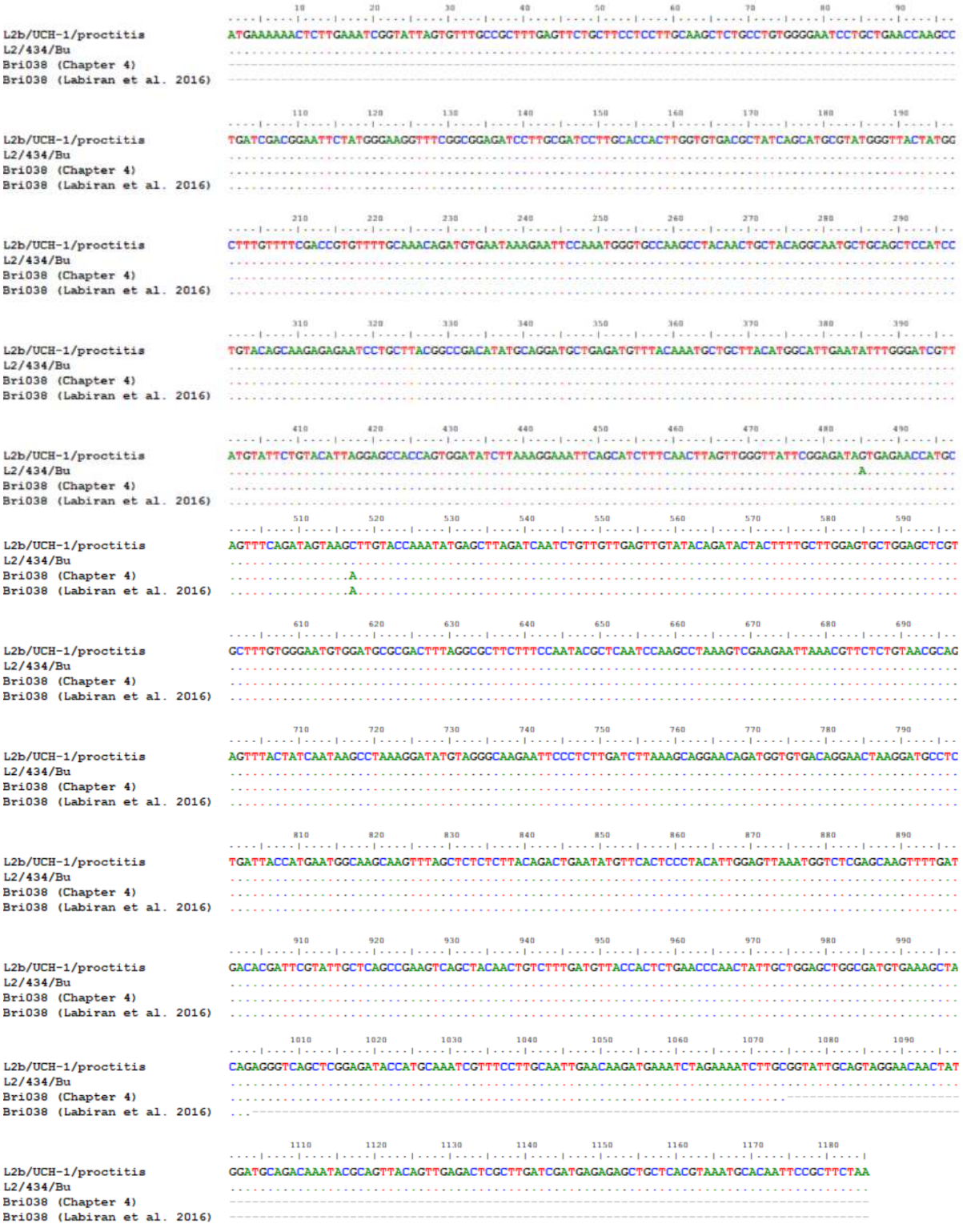

570 580 590 600 610 620 630
C/TW3 (NC_023060) ATCAAAGAAGAATTTCGAAAAGTCAGAGAACATTTTGATCTTCCAATTGTAGATAGAAGAGATATTTGTG
W1
W2
W3
W4
W5
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808)
Ia/SotonIa3 (HE601809)
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)


640 650 660 670 680 690 700
C/TW3 (NC_023060) ATAAAAAAGAAGCTGCACATGTGCTGAATTATTCAGATGGTTTCATTGTGAAAACAGCGTTTGTTCATCA
W1
W2
W3
W4
W5 T
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808) C
Ia/SotonIa3 (HE601809) C
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952)
L1/224 (HE601953)
L1/440 (HE601950)
L1/1322/p2 (HE601951)
L2/434/Bu (AM884176)
L2/25667R (HE601954)
L2b/Ams1 (HE601959)
L2b/Ams2 (HE601961)
L2b/Ams4 (HE601964)
L2b/Ams5 (HE601965)
L2b/Canada1 (HE601963)
L2b/Canada2 (HE601957)
L2b/UCH-1/proctitis (AM884177)
L2b/UCH-2 (HE601956)
L2b/8200/07 (HE601795)
L2b/CV204 (HE601960)
L2b/795 (HE601949)
L2b/LST (HE601958)
L3/404/LN (HE601955)


710 720 730 740 750 760
C/TW3 (NC_023060) GACAACAATGGATTCTTCGGTAGAGACTCTGACTGCACTTGCACAAACAGTTATTCCTGGATAA
W1
W2
W3
W4
W5 A
W6
A/363 (HE601796)
A/2497 (NC_017437)
A/5291 (HE601810)
A/Har13 (NC_007429)
A/7249 (HE601797)
B/TZ1A828-OT (NC_012687)
B/Jali20 (NC_012686)
D/UW3/CX (NC_000117)
D/SotonD1 (HE601798)
D/SotonD5 (HE601799)
D/SotonD6 (HE601800)
D(s)/2923 (NZ_ACFJ01000001)
E/SotonE8 (HE601803)
E/SotonE4 (HE601802)
E/Bour (HE601870)
E/150 (NC_017439)
E/SW2 (NC_017441)
E/SW3 (HE601801)
F/SotonF3 (HE601806)
F/SW4 (HE601804)
F/SW5 (HE601805)
G/11074 (CP001889)
G/11222 (CP001888)
G/SotonG1 (HE601807)
G/9768 (NC_017429)
G/9301 (NC_017432)
Ia/SotonIa1 (HE601808)
Ia/SotonIa3 (HE601809)
J/UW36 (AY096814)
K/SotonK1 (HE601794)
L1/115 (HE601952) A
L1/224 (HE601953) A
L1/440 (HE601950) A
L1/1322/p2 (HE601951) A
L2/434/Bu (AM884176) A
L2/25667R (HE601954) A
L2b/Ams1 (HE601959) A
L2b/Ams2 (HE601961) A
L2b/Ams4 (HE601964) A
L2b/Ams5 (HE601965) A
L2b/Canada1 (HE601963) A
L2b/Canada2 (HE601957) A
L2b/UCH-1/proctitis (AM884177) A
L2b/UCH-2 (HE601956) A
L2b/8200/07 (HE601795) A
L2b/CV204 (HE601960) A
L2b/795 (HE601949) A
L2b/LST (HE601958) A
L3/404/LN (HE601955) A

APPENDIX 3 Alignment of *ompA* from Bri038

The *ompA* sequence of Bri038 obtained in Chapter 4 was aligned against the *ompA* of Bri038 previously obtained by Labiran *et al.* (2016). Also included in the sequence alignment were the *ompA* sequences of L2b/UCH-1/proctitis (Genbank Accession No. AM884177) and L2/434/Bu (Genbank Accession No. AM884176). Nucleotide positions are given relative to L2b/UCH-1/proctitis. Coloured dots indicate bases that were identical to L2b/UCH-1/proctitis at that nucleotide position. The C517A SNP is visible in both Bri038 *ompA* sequences.



APPENDIX 4 Publication Two: MLVA-*ompA* Genotyping Survey of LGV Strains of *C. trachomatis* in London

PLOS ONE

RESEARCH ARTICLE

High-resolution genotyping of Lymphogranuloma Venereum (LGV) strains of *Chlamydia trachomatis* in London using multi-locus VNTR analysis-*ompA* genotyping (MLVA-*ompA*)

Chloe Manning[1☯], Colette O'Neill[1☯], Ian N. Clarke[1☯], Monica Rebec[2‡], Penelope R. Cliff[3‡], Peter Marsh[4☯]

1 Department of Molecular Microbiology, Division of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **2** Department of Microbiology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, London, United Kingdom, **3** Department of Infection Sciences, St Thomas' Hospital, London, United Kingdom, **4** Public Health England, Porton Down, Salisbury, United Kingdom

☯ These authors contributed equally to this work.
‡ MR and PRC also contributed equally to this work.
* c.e.manning@soton.ac.uk

OPEN ACCESS

Citation: Manning C, O'Neill C, Clarke IN, Rebec M, Cliff PR, Marsh P (2021) High-resolution genotyping of Lymphogranuloma Venereum (LGV) strains of *Chlamydia trachomatis* in London using multi-locus VNTR analysis-*ompA* genotyping (MLVA-*ompA*). PLoS ONE 16(7): e0254233. https://doi.org/10.1371/journal.pone.0254233

Editor: Sylvia Maria Bruisten, GGD Amsterdam, NETHERLANDS

Received: January 4, 2021

Accepted: June 22, 2021

Published: July 8, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files.

Funding: This work was supported by a Public Health England PhD Studentship for Chloe Manning (PhD2017/33) and a University of Southampton Vice-Chancellor's Scholarship (FACVC17A/Clarke/201718/264). The funders had no role in study design, data collection and

Abstract

Background

Lymphogranuloma venereum (LGV) is caused by *Chlamydia trachomatis* strains with *ompA* genotypes L1 to L3. An LGV epidemic associated with the L2b genotype has emerged in the past few decades amongst men who have sex with men (MSM). *C. trachomatis* genotypes can be discriminated by outer membrane protein A gene (*ompA*) sequencing, however this method has limited resolution. This study employed a high-resolution genotyping method, namely, multi-locus tandem repeat (VNTR) analysis with *ompA* sequencing (MLVA-*ompA*), to assess the distribution of LGV MLVA-*ompA* genotypes amongst individuals attending genitourinary medicine (GUM) clinics in London.

Methods

Clinical specimens were collected from individuals attending eight London-based GUM clinics. Specimens that tested positive for *C. trachomatis* by commercial nucleic acid amplification test (NAAT) were confirmed as LGV by *pmpH* real-time PCR. LGV-positive DNA extracts were subsequently genotyped using MLVA-*ompA*.

Results

Two hundred and thirty DNA extracts were confirmed as LGV, and 162 (70%) yielded complete MLVA-*ompA* genotypes. Six LGV MLVA-*ompA* genotypes were identified: 1.9.2b-L2, 1.9.3b-L2b, 1.9.2b-L2b, 1.9.2b-L2b/D, 1.4a.2b-L2b, and 5.9.2b-L1. The following LGV *ompA* genotypes were identified (in descending order of abundance): L2, L2b, L2b/D, and

analysis, decision to publish, or preparation of the manuscript.

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

L1. Eight *ompA* sequences with the hybrid L2b/D profile were detected. The hybrid sequence was identical to the *ompA* of a recombinant L2b/D strain detected in Portugal in 2017.

Conclusions

The L2 *ompA* genotype was found to predominate in the London study population. The study detected an unusual hybrid L2b/D *ompA* profile that was previously reported in Portugal. We recommend further monitoring and surveillance of LGV strains within the UK population.

Introduction

Lymphogranuloma Venereum (LGV) is caused by *Chlamydia trachomatis ompA* genotypes L1, L2 and L3 [1]. In 2003, an outbreak of LGV was reported in a sexual network of men who have sex with men (MSM) presenting with symptoms of proctitis in Rotterdam [2]. Sequencing of the *ompA* gene demonstrated that the initial outbreak was caused by a new genetic variant designated L2b [3]. This variant has subsequently been implicated in LGV outbreaks worldwide [4–10]. The United Kingdom has the highest number of confirmed LGV cases in Europe, with a total of 6,752 UK LGV diagnoses made between 2003 and 2018 [11, 12]. A report showed that 617/919 (67%) of UK LGV diagnoses were made in London [11].

Whole genome sequencing has proven informative for elucidating the evolutionary history and diversity of *C. trachomatis* [13, 14], however the high resolution provided by this method is not always necessary [15]. For decades, sequence analysis of the *ompA* gene was used to differentiate *C. trachomatis* strains [16]. However, the *ompA* gene is not always an accurate epidemiological marker when used on its own, with *ompA* shown to be a recombination hotspot in the genome [13]. *OmpA* genotyping has been largely superseded by more discriminatory genotyping systems including multi-locus sequence typing (MLST) [15, 17], and multi-locus variable number tandem repeat (VNTR) analysis with *ompA* genotyping (MLVA-*ompA*) [18].

MLVA-*ompA* targets variation in the number of repeating mononucleotides at three VNTR loci dispersed throughout the chlamydial genome (i.e. CT1335, CT1299, and CT1291), coupled with sequencing of the *ompA* gene [18]. The MLVA-*ompA* system has been successfully applied to genotype *C. trachomatis* strains globally [19–23]. MLVA-*ompA* has a high degree of resolution which is essential for isolate discrimination, with earlier studies reporting a discriminatory index between 0.94 and 0.99 [18, 24], as measured by Simpson's Index of Diversity [25]. Whilst MLVA-*ompA* has not previously been applied on a large-scale to genotype LGV strains in the United Kingdom; LGV clinical samples have been genotyped successfully using the system in Brighton (n = 11) [19] and Southampton (n = 1) [24]. The study in Brighton identified nine distinct LGV MLVA-*ompA* genotypes, with seven genotypes detected within isolates assigned an *ompA* genotype L2b [19]. The genotypic diversity exhibited within LGV strains in Brighton raised questions about whether a similar extent of genotypic diversity might exist within LGV strains circulating in other UK cities. Given the high prevalence of LGV in London and its substantial MSM population [11], it was decided to apply the MLVA-*ompA* system to genotype LGV clinical DNA extracts from this region. The aim of this study was to assess the distribution of LGV MLVA-*ompA* genotypes from clinical specimens sourced from individuals attending eight London-based GUM clinics.

Methods

Clinical specimens

Clinical specimens including rectal swabs, throat swabs, urine and "pooled 3-in-1" specimens (a rectal swab and pharyngeal swab in urine, from a single patient) were collected from patients attending eight London-based GUM clinics. Specimens collected from 56 Dean Street, John Hunter Clinic, 10 Hammersmith Broadway, and Jefferiss Wing were tested for *C. trachomatis* at North West London Partnership (NWLP), hosted by Imperial College Healthcare NHS Trust; and specimens collected from Burrell Street Clinic, Streatham Hill Clinic, Walworth Road Clinic, and Harrison Wing were tested for *C. trachomatis* at Viapath, a private-sector diagnostic laboratory based at St Thomas' Hospital.

At Viapath, the Aptima CT/NG Combo 2 Assay (Hologic®, US) was used for *C. trachomatis* detection. NAAT-positive specimens from MSM were reflex tested for LGV at Viapath. Nucleic acids were extracted using the Complex 200 Protocol for the QIAsymphony DSP Virus/Pathogen kit (QIAGEN, US), followed by an in-house triplex LGV PCR assay targeting the *pmpH* gene located on the *C. trachomatis* chromosome; in addition to an 88-base pair region of the *C. trachomatis* cryptic plasmid; and the human *RNase P* gene as an internal control [26].

For the majority of clinics that send specimens to NWLP for *C. trachomatis* testing, there is a reflex to LGV test all rectal NAAT-positive specimens. At NWLP, the BD ProbeTec™ CT/GC Amplified DNA Assay (Becton-Dickinson, US) was used for *C. trachomatis* detection. Purified nucleic acids were isolated from each specimen using the MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche Life Science, UK) before application of the triplex LGV PCR assay (as described for Viapath) [26].

For this study, DNA extracts that were confirmed as LGV at NWLP and Viapath were selected consecutively with no bias for positivity strength. DNA extracts included in this study were collected from patients between February 2018 and June 2019. DNA extracts were transported on dry ice to the University of Southampton Molecular Microbiology Group for genotyping. All patient-identifiable information was removed from each DNA extract prior to dispatch to Southampton. Extracts were subsequently stored at -20˚C.

PCR amplification of VNTR and *ompA* sequences

VNTR and *ompA* sequences were amplified from the DNA extracts using PCR according to Wang *et al* [22]. For *ompA*, a fragment of this gene (ca 1,000bp) was amplified using primers PCTM3 and NR1 [27], whilst the three VNTR regions were amplified using primers described by Pedersen *et al* [18]. Extracts that did not produce VNTR amplicons using these primers were amplified using primers CT1335F* and CT1335R*, CT1299F* and CT1299R*, and CT1291F* and CT1291R* [28] (S1 Table). The forward primers annealed upstream of the original VNTR amplicon sequences, and the reverse primers downstream, so that the original amplicon sequences were encompassed by the alternative primers. PCR reactions were carried out in 20 μL volumes consisting of: 10 μL Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific™, UK), 0.5μM of the forward and reverse primers (Eurogentec, Belgium), and 1 μL of DNA. PCR products were loaded onto 2% (w/v) agarose gels for the purpose of checking amplicon size and quality. The amplicons were subsequently purified using the Wizard SV Gel and PCR Clean-Up System (Promega, UK) for sequencing. PCR amplicons were commercially sequenced at Source Bioscience (Cambridge, UK).

DNA sequence analysis of MLVA-*ompA* markers

Alphabetical *ompA* genotypes were assigned to each extract via *ompA* sequence comparison to the NCBI database using Basic Local Alignment Search Tool (BLAST) [22]. VNTR sequences were compared to those described in Pedersen *et al* [18], Wang *et al* [22], Labiran *et al* [24], and Satoh *et al* [21]; and a single-digit number assigned at each VNTR locus based on the number of repeating mononucleotides [22]. The final MLVA-*ompA* genotype was designated by a three-digit code in the order: CT1335, CT1299, and CT1291, followed by the alphabetical *ompA* genotype (e.g. 3.9.3-G).

Bioinformatics

To confirm the VNTR profile of the Portuguese strain (strain Ct_L2b/D_PT05; European Nucleotide Archive (ENA) accession number CAAKND010000000) [29] *in silico*,. fastq files were downloaded and converted to .fasta format using the conversion tool on the public server at usegalaxy.org. The resulting .fasta files were opened using the BioEdit (version 7.0.5.3) alignment software, and the VNTRs were located by inputting the VNTR primer sequences into the search tool.

Ethics

This study was approved by the East of Scotland Research Ethics Committee (REC reference 19/ES/0012). All DNA extracts were received anonymised and unlinked. No patient clinical data, including age, sex, risk behaviour and clinical symptoms, was collected for this study. The requirement for informed consent was waived by the ethics committee.

Results

Clinical DNA extracts

A total of 230 DNA extracts that were confirmed as LGV using the *pmpH* real-time PCR-based assay were obtained for this study. These included 180 DNA extracts from NWLP and 50 DNA extracts from Viapath. Extracts from NWLP were from rectal swabs, and extracts from Viapath were from 23 pooled "3-in-1" specimens, 24 rectal swabs, 1 throat swab and 2 urines.

OmpA genotypes identified in this study

One hundred and seventy three extracts (75.2%) were assigned an *ompA* genotype in this study (S2 Table). Of these, 164 extracts (94.8%) were assigned an LGV *ompA* genotype, and 9 were assigned non-LGV *ompA* genotypes. The most prevalent LGV *ompA* genotype identified in the study was L2 (n = 81, 49.3%), followed by L2b (n = 72, 43.9%). The non-LGV *ompA* genotypes identified were genotypes E (n = 3), G (n = 5), and J (n = 1).

MLVA-*ompA* genotypes identified in this study

Sequence data were obtained for all four loci for 162/230 (70.4%) of the extracts (S2 Table). Of these, 159 (98.1%) were assigned LGV MLVA-*ompA* genotypes, and the remaining 3 extracts were assigned 3.9.3-G. Six distinct LGV MLVA-*ompA* genotypes were identified in this study (Table 1). The most prevalent LGV MLVA-*ompA* genotypes were 1.9.2b-L2 (n = 78, 49.1%), 1.9.2b-L2b (n = 53, 33.3%), and 1.9.3b-L2b (n = 16, 10.1%). Also detected were 5.9.2b-L1 (n = 3, 1.9%), 1.4a.2b-L2b (n = 1, 0.6%), and 1.9.2b-L2b/D (n = 8, 5.0%).

VNTR sequence variants identified in the study. The VNTR variant code, CT1291 type 3b (AAAATAGTCTA-**9C**-TATTG), was identified in 20 extracts in this study. Sixteen extracts

Table 1. Complete LGV MLVA-*ompA* genotypes identified in this study (n = 159).

ompA	MLVA†	n (% of 159 extracts)
L1	5.9.2b	3 (1.9)
L2	1.9.2b	78 (49.1)
L2b	1.4a.2b	1 (0.6)
	1.9.2b	53 (33.3)
	1.9.3b	16 (10.1)
L2b/D	1.9.2b	8 (5.0)

† MLVA genotype was designated by the 3 VNTR loci in the order: CT1335; CT1299; and CT1291.

Extracts with partial MLVA-*ompA* profiles were excluded from this table.

https://doi.org/10.1371/journal.pone.0254233.t001

with the 3b variant code could be assigned an *ompA* genotype, of which all were *ompA* genotype L2b (S2 and S3 Tables). The complete MLVA-*ompA* genotype of these sixteen extracts was 1.9.3b-L2b. The CT1291 type 3b was previously identified by Satoh *et al* [21] and assigned to the reference strains L1/440/Bu and L2/434/Bu.

VNTR variants identified by Pedersen *et al* [18] and Wang *et al* [22] were detected in this study (S3 Table). The extract assigned a CT1299 type 4a in this study was the 4a variant (TTTTTATTCT-**10C**-T3C-ATCAAA) first identified in Wang *et al* [22]. All extracts assigned a CT1291 type 2b in this study were the 2b variant (AAAATAGTCTA-**8C**-TATTG) initially identified in Wang *et al* [22] (S3 Table).

In our study, CT1299 VNTR type 9 (TTTTTATTCT-3C2T-**6C**-ATCAAA) was assigned to 161/164 (98.2%) extracts with LGV *ompA* genotypes (S3 Table).

Detection of a hybrid L2b/D *ompA* genotype. We identified eight *ompA* sequences with an L2b/D hybrid profile (S1 Fig), i.e. whilst the first 365bp (numbers given relative to L2b/UCH-1/proctitis) of each sequence was identical to *ompA* L2 and L2b reference sequences (L2/434/Bu and L2b/UCH-1/proctitis); the region spanning 366bp-1,023bp revealed an *ompA* genotype D profile matching the reference strain D/UW-3/CX (Genbank accession no. NC_000117.1). Nucleotide BLAST of the eight sequences indicated 100% sequence identity to the *ompA* sequence of a novel hybrid L2b/D strain identified in Portugal [29] (Genbank accession no. MN094864.1). The eight hybrid sequences were designated L2b/D, to distinguish them from extracts with an *ompA* sequence matching L2b/UCH-1/proctitis.

All extracts assigned the L2b/D *ompA* genotype could be assigned a full MLVA-*ompA* genotype, all eight of which were 1.9.2b-L2b/D (S2 Table). We confirmed the VNTR profile of the Portuguese strain (ENA accession number CAAKND010000000) *in silico*: GAAAAAG-**9T8A**-GCTTTTGT at CT1335 (CT1335 type 1), TTTTTATTCT-3C2T-**6C**-ATCAAA at CT1299 (CT1299 type 9), and AAAATAGTCTA-**8C**-TATTG at CT1291 (CT1291 type 2b), corresponding to the same MLVA-*ompA* genotype of 1.9.2b-L2b/D as identified in this study.

Discussion

This study represents the first MLVA-*ompA* genotyping survey of LGV strains of *C. trachomatis* in a London population. Amongst the extracts that could be assigned a type at all four loci, we identified six distinct LGV MLVA-*ompA* genotypes: 5.9.2b-L1, 1.9.2b-L2, 1.9.2b-L2b, 1.9.3b-L2b, 1.4a.2b-L2b, and 1.9.2b-L2b/D.

In accordance with a previous study [30], we found that the L2b *ompA* variant was not the most common in the pooled data set; the L2 *ompA* sequence predominated. However, in the previous study this varied by country, and the L2 *ompA* sequence predominated in Austria

and Croatia, whilst the L2b *ompA* sequence predominated in the UK. Whilst the predominance of the L2 *ompA* genotype has been documented in France [31], Sweden [5], Spain [32], and Austria and Croatia [30], this is the first report of its predominance within the United Kingdom. A retrospective study of LGV *C. trachomatis* strains collected in France between 2010 and 2015 by Peuchant *et al.* [31] demonstrated that the proportion of LGV cases caused by L2b declined after 2012, whilst the proportion caused by *ompA* genotype L2 increased from 2012 onwards, which supports our findings. Our study showed that of those L2 extracts with a complete MLVA-*ompA* genotype (78/81, 96.3%), 100% had the MLVA-*ompA* genotype 1.9.2b-L2. This VNTR profile, 1.9.2b, was found to be shared amongst extracts designated L2, L2b and L2b/D in this study. These data show that at least three LGV *ompA* genotypes with the same VNTR profile are co-circulating within the London population. This is likely due to recombination within the *ompA* gene, given that *ompA* is known to be a recombination hotspot within the chlamydial genome [13, 33].

We noted the most diversity in MLVA-*ompA* genotypes within extracts designated *ompA* genotype L2b, with VNTR profiles 1.9.2b, 1.9.3b, and 1.4a.2b assigned. Diversity in L2b genotypes has previously been reported in Brighton [19]. Interestingly, whilst 1.9.2b-L2b and 1.9.3b-L2b comprised 53/159 (33.3%) and 16/159 (10.1%) respectively of all extracts assigned a complete LGV MLVA-*ompA* genotype in this study (Table 1); the Brighton study reported only three LGV cases with 1.9.2b-L2b and none with 1.9.3b-L2b. Further, the Brighton study identified eight LGV MLVA-*ompA* genotypes that were not detected in our London study population. These regional differences in MLVA-*ompA* genotypes between Brighton and London are likely the result of distinct dissemination patterns within each population; however, given that the Brighton study took place between 2011 and 2013, and the specimens for this London study were collected between 2018 and 2019; these differences could represent a temporal shift in genotypes.

We detected a hybrid L2b/D *ompA* sequence in DNA extracts from Viapath (n = 1) and NWLP (n = 7). All were assigned the MLVA-*ompA* genotype 1.9.2b-L2b/D. The hybrid *ompA* sequence was identical to the *ompA* of a recombinant L2b/D strain detected in Portugal [29] (S1 Fig). As of 2019, a total of 25 cases of the recombinant L2b/D strain have been reported in Portugal. Our study is the first report of the hybrid L2b/D *ompA* sequence in the United Kingdom. We demonstrated by *in silico* means that the Portuguese L2b/D strain possessed the 1.9.2b-L2b/D MLVA-*ompA* genotype. This result confirmed that the Portuguese L2b/D strain and the eight extracts assigned L2b/D from our London study population shared the same MLVA-*ompA* genotype. Whole genome sequencing of the Portuguese strain performed by Borges *et al.* [29] revealed that the strain resulted from the transfer of a 4.2kbp fragment from a *C. trachomatis* D strain to an L2b strain. The recombinant fragment comprised 75% of the *ompA* gene encoding the major outer membrane protein (MOMP), and four genes downstream of *ompA* each with functional roles in protein synthesis. This widespread genetic recombination, particularly within the MOMP epitope region that is responsible for influencing the ability of *C. trachomatis* strains to interact with the host immune response [34, 35], may have implications for the transmission and pathogenic capability of the hybrid strain [36, 37]. A limitation of our study was that we did not collect clinical data relating to patient symptoms, and as a result, we are unable to comment on the clinical presentation of patients with the L2b/D *ompA* sequence in the London study population. However it was noted by Borges *et al.* [29] that all of the individuals infected with the hybrid L2b/D strain presented with similar symptoms and clinical features (i.e. rectal pain, anal discharge and rectal bleeding), that are consistent with a typical LGV infection [1]. Of note, many of the individuals infected with the L2b/D strain in Portugal were involved in international sexual networks, which would explain how the variant likely reached our study population.

We detected the CT1291 variant code, type 3b (AAAATAGTCTA-**9C**-TATTG) (S3 Table) in 20 extracts. Sixteen of these extracts could be assigned a complete MLVA-*ompA* genotype, of which all were 1.9.3b-L2b. Prior to this study, CT1291 type 3b was detected by Satoh *et al*. in the L1/440/Bu and L2/434/Bu reference strains [21], and these were assigned the MLVA-*ompA* genotypes 5.9.3b-L1 and 5.9.3b-L2. The Satoh *et al*. study did not include any L2b isolates. The CT1291 type 3b was not detected in any of the 44 clinical isolates also MLVA-*ompA* genotyped by Satoh *et al*. In this study, we also detected VNTR variant codes first identified in Pedersen *et al* [18] and Wang *et al* [22], including CT1299 type 4a (TTTTTATTCT-**10C**-T3C-ATCAAA), CT1299 type 9 (TTTTTATTCT-3C2T-**6C**-ATCAAA), and CT1291 type 2b (AAAATAGTCTA-**8C**-TATTG).

The MLVA-*ompA* genotyping method had limited resolution when applied to our London study population, with three LGV MLVA-*ompA* genotypes (1.9.2b-L2, 1.9.3b-L2b, and 1.9.2b-L2b) comprising 92.5% of extracts assigned complete MLVA-*ompA* genotypes in the study. As a result, we did not reach a Simpson's index of diversity of 0.95, the index value for a genotyping system to be considered to have more or less "ideal" resolution [38].

We detected the non-LGV *ompA* genotypes E (n = 3), G (n = 5) and J (n = 1) in this study. Genotypes G and J have commonly been found in the rectum of MSM [39, 40], and co-infections of LGV and urogenital genotype E infection have been reported previously [8]. However, extracts that were genotyped in this study were those that had given a positive result in the LGV biovar assay at NWLP and Viapath; that is, only extracts with the 36-bp deletion within the *pmpH* gene that is characteristic of LGV strains. Given the low prevalence of these non-LGV *ompA* genotypes in our study (9/173, 5.2%, of those with an assigned *ompA* genotype), there are a few possible explanations for their detection. Firstly, the nine extracts could have been false positives of the LGV biovar assay. The likelihood of this is slight—the assay has demonstrated excellent diagnostic performance in differentiating LGV and non-LGV infections in previous studies [41]. It is more likely that the individuals with a non-LGV *ompA* genotype in this study were infected with both an LGV strain and a non-LGV strain of *C. trachomatis*. Another possible explanation is that mixed infection could have resulted in *pmpH* variants caused by genetic exchange between LGV and genotype G, J or E *C. trachomatis* strains [8]. Given that the evolution of *C. trachomatis* is mainly driven by recombination [42], and many studies have reported co-infections with LGV strains and non-LGV *C. trachomatis* strains with invasive and non-invasive urogenital *ompA* genotypes [8, 42, 43], this explanation is still plausible. Mixed infections can help to facilitate the selection of new recombinants, such as the hybrid Portuguese strain [29], and the L2c strain described by Somboona *et al* [44], that were both caused by unique recombination events between L2b (and L2, respectively) and D genotypes.

Conclusions

In conclusion, we have demonstrated that the predominant *ompA* genotype within our London study population is the L2 *ompA* genotype, and not the L2b *ompA* genotype that has been reported in UK populations since 2005 [45]. We provide the first UK report of a hybrid L2b/D *ompA* profile previously detected in Portugal. These findings highlight the ever-changing nature of the LGV epidemic, and we urge for attentive LGV surveillance strategies to continue.

Supporting information

S1 Table. Primer sequences for PCR of MLVA-*ompA* markers.
(DOCX)

S2 Table. Distribution of MLVA-*ompA* genotypes from clinical DNA extracts sourced from individuals attending eight London-based GUM clinics (n = 230).
(XLSX)

S3 Table. VNTR sequence analysis of extracts assigned LGV *ompA* genotypes (n = 164). VNTR type codes of extracts that were assigned non-LGV *ompA* genotypes (n = 9) or that could not be assigned an *ompA* genotype (n = 57), were excluded from this table.
(DOCX)

S1 Fig. Alignment of partial *ompA* gene from extracts possessing a hybrid L2b/D-Da *ompA* profile (n = 8). The *ompA* sequence of L2b/UCH-1/proctitis (Genbank accession no. AM884177.1) was used as a reference. Nucleotide numbers are given according to L2b/UCH-1/proctitis. L2b/D refers to the *ompA* sequence of the Portuguese L2b/D strain (Genbank accession no. MN094864.1). The *ompA* sequences of L2/434/Bu (Genbank accession no. AM884176.1) and D/UW-3 (Genbank accession no. NC_000117.1) were included in the alignment. Coloured dots indicate bases that matched L2b/UCH-1/proctitis.
(DOCX)

Acknowledgments

We wish to gratefully acknowledge Dr Achyuta Nori at Burrell Street Clinic for his support and for providing guidance on the project, and to the Chlamydia Biobank (http://www.chlamydiabiobank.co.uk/) for providing us with genomic DNA extracted from the SotonK1 strain (Genbank accession no. NC_020965) for use as a positive control.

Author Contributions

Conceptualization: Chloe Manning, Ian N. Clarke, Peter Marsh.

Formal analysis: Chloe Manning.

Funding acquisition: Ian N. Clarke, Peter Marsh.

Investigation: Chloe Manning.

Methodology: Chloe Manning, Monica Rebec, Penelope R. Cliff, Peter Marsh.

Project administration: Chloe Manning.

Resources: Monica Rebec, Penelope R. Cliff.

Supervision: Colette O'Neill, Ian N. Clarke, Peter Marsh.

Writing – original draft: Chloe Manning.

Writing – review & editing: Chloe Manning, Colette O'Neill, Ian N. Clarke, Monica Rebec, Penelope R. Cliff, Peter Marsh.

References

1. de Vries HJC, de Barbeyrac B, de Vrieze NHN, Viset J., White JA, Vall-Mayans M, et al. 2019 European guideline on the management of lymphogranuloma venereum. J Eur Acad Dermatology Venereol. 2019 Oct 26; 33(10):1821–8. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/jdv.15729 PMID: 31243838
2. Götz HM, Ossewaarde JM, Nieuwenhuis RF, van der Meijden WI, Dees J, Thio B, et al. [A cluster of lymphogranuloma venereum among homosexual men in Rotterdam with implications for other countries in Western Europe]. Ned Tijdschr Geneeskd. 2004 Feb 28; 148(9):441–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15038207

3. Spaargaren J, Fennema HSA, Morré SA, de Vries HJC, Coutinho RA. New lymphogranuloma venereum Chlamydia trachomatis variant, Amsterdam. Emerg Infect Dis. 2005 Jul; 11(7):1090–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16022786 https://doi.org/10.3201/eid1107.040883

4. Spaargaren J, Schachter J, Moncada J, de Vries HJC, Fennema HSA, Peña AS, et al. Slow Epidemic of Lymphogranuloma Venereum L2b Strain. Emerg Infect Dis. 2005 Nov; 11(11):1787–8. Available from: http://wwwnc.cdc.gov/eid/article/11/11/05-0821_article.htm PMID: 16318741

5. Isaksson J, Carlsson O, Airell Å, Strömdahl S, Bratt G, Herrmann B. Lymphogranuloma venereum rates increased and Chlamydia trachomatis genotypes changed among men who have sex with men in Sweden 2004–2016. J Med Microbiol. 2017 Nov; 66(11):1684–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29022544 https://doi.org/10.1099/jmm.0.000597

6. Nieuwenhuis RF, Ossewaarde JM, Götz HM, Dees J, Thio HB, Thomeer MGJ, et al. Resurgence of lymphogranuloma venereum in Western Europe: an outbreak of Chlamydia trachomatis serovar l2 proctitis in The Netherlands among men who have sex with men. Clin Infect Dis. 2004 Oct 1; 39(7):996–1003. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15472852 https://doi.org/10.1086/423966

7. Ward H, Martin I, Macdonald N, Alexander S, Simms I, Fenton K, et al. Lymphogranuloma venereum in the United kingdom. Clin Infect Dis. 2007 Jan 1; 44(1):26–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17143811 https://doi.org/10.1086/509922

8. Rodriguez-Dominguez M, Gonzalez-Alba JM, Puerta T, Menendez B, Sanchez-Diaz AM, Canton R, et al. High Prevalence of Co-Infections by Invasive and Non-Invasive Chlamydia trachomatis Genotypes during the Lymphogranuloma Venereum Outbreak in Spain. Dean D, editor. PLoS One. 2015 May 12; 10(5):e0126145. Available from: https://dx.plos.org/10.1371/journal.pone.0126145 PMID: 25965545

9. Kropp RY, Wong T, Canadian LGV Working Group. Emergence of lymphogranuloma venereum in Canada. CMAJ. 2005 Jun 21; 172(13):1674–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15928239 https://doi.org/10.1503/cmaj.050621

10. Stark D, van Hal S, Hillman R, Harkness J, Marriott D. Lymphogranuloma Venereum in Australia: Anorectal Chlamydia trachomatis Serovar L2b in Men Who Have Sex with Men. J Clin Microbiol. 2007 Mar 1; 45(3):1029–31. Available from: https://jcm.asm.org/content/45/3/1029 PMID: 17251405

11. Public Health England. Recent trends in Lymphogranuloma Venereum (LGV) diagnoses in the United Kingdom: 2003–2016. 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/653619/hpr3717_LGV.pdf

12. European Centre for Disease Control. Lymphogranuloma venereum: Annual Epidemiological Report for 2018. 2020.

13. Harris SR, Clarke IN, Seth-Smith HMB, Solomon AW, Cutcliffe LT, Marsh P, et al. Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet. 2012 Mar 11; 44(4):413–9, S1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22406642 https://doi.org/10.1038/ng.2214

14. Seth-Smith HMB, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, et al. Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture. Genome Res. 2013 May 1; 23(5):855–66. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.150037.112 PMID: 23525359

15. Herrmann B, Isaksson J, Ryberg M, Tångrot J, Saleh I, Versteeg B, et al. Global Multilocus Sequence Type Analysis of Chlamydia trachomatis Strains from 16 Countries. J Clin Microbiol. 2015 Jul; 53(7):2172–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25926497 https://doi.org/10.1128/JCM.00249-15

16. Dean D, Schachter J, Dawson CR, Stephens RS. Comparison of the Major Outer Membrane Protein Variant Sequence Regions of B/Ba Isolates: A Molecular Epidemiologic Approach to Chlamydia trachomatis Infections. J Infect Dis. 1992 Aug 1; 166(2):383–92. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/166.2.383 PMID: 1634810

17. Klint M, Fuxelius H-H, Goldkuhl RR, Skarin H, Rutemark C, Andersson SGE, et al. High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis. J Clin Microbiol. 2007 May; 45(5):1410–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17329456 https://doi.org/10.1128/JCM.02301-06

18. Pedersen LN, Pødenphant L, Møller JK. Highly discriminative genotyping of Chlamydia trachomatis using omp1 and a set of variable number tandem repeats. Clin Microbiol Infect. 2008 Jul; 14(7):644–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18558936 https://doi.org/10.1111/j.1469-0691.2008.02011.x

19. Labiran C, Marsh P, Zhou J, Bannister A, Clarke IN, Goubet S, et al. Highly diverse MLVA-ompA genotypes of rectal Chlamydia trachomatis among men who have sex with men in Brighton, UK and

evidence for an HIV-related sexual network. Sex Transm Infect. 2016; 92(4):299–304. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26487754 https://doi.org/10.1136/sextrans-2015-052261

20. Peuchant O, Le Roy C, Herrmann B, Clerc M, Bébéar C, de Barbeyrac B. MLVA subtyping of genovar E Chlamydia trachomatis individualizes the Swedish variant and anorectal isolates from men who have sex with men. PLoS One. 2012; 7(2):e31538. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22363667 https://doi.org/10.1371/journal.pone.0031538

21. Satoh M, Ogawa M, Saijo M, Ando S. Multilocus VNTR analysis-ompA typing of venereal isolates of Chlamydia trachomatis in Japan. J Infect Chemother. 2014 Oct; 20(10):656–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25066435 https://doi.org/10.1016/j.jiac.2014.06.010

22. Wang Y, Skilton RJ, Cutcliffe LT, Andrews E, Clarke IN, Marsh P. Evaluation of a high resolution genotyping method for Chlamydia trachomatis using routine clinical samples. PLoS One. 2011 Feb 11; 6(2): e16971. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21347295 https://doi.org/10.1371/journal.pone.0016971

23. Bom RJM, Christerson L, Schim van der Loeff MF, Coutinho RA, Herrmann B, Bruisten SM. Evaluation of high-resolution typing methods for Chlamydia trachomatis in samples from heterosexual couples. J Clin Microbiol. 2011 Aug; 49(8):2844–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21653758 https://doi.org/10.1128/JCM.00128-11

24. Labiran C, Rowen D, Clarke IN, Marsh P. Detailed molecular epidemiology of Chlamydia trachomatis in the population of Southampton attending the genitourinary medicine clinic in 2012–13 reveals the presence of long established genotypes and transitory sexual networks. PLoS One. 2017; 12(9):e0185059. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28945766 https://doi.org/10.1371/journal.pone.0185059

25. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988 Nov; 26(11):2465–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3069867 https://doi.org/10.1128/jcm.26.11.2465-2466.1988

26. Mayes J, Liyanarachchy N, Balm MND. Nucleic Acid Extraction from Residual Aptima Clinical Samples: Evaluation of Automated Platform To Enhance Workflow. Ledeboer NA, editor. J Clin Microbiol. 2017 Jun; 55(6):1968–9. Available from: https://jcm.asm.org/content/55/6/1968 PMID: 28404675

27. Lan J, Ossewaarde JM, Walboomers JM, Meijer CJ, van den Brule AJ. Improved PCR sensitivity for direct genotyping of Chlamydia trachomatis serovars by using a nested PCR. J Clin Microbiol. 1994 Feb; 32(2):528–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8150967 https://doi.org/10.1128/jcm.32.2.528-530.1994

28. Labiran C. Molecular epidemiology of Chlamydia trachomatis: Evaluation, implementation and development of high resolution genotyping. 2014. https://eprints.soton.ac.uk/374912/1/Clare%2520Labiran_%2520Thesis_%252010_June_14.pdf

29. Borges V, Cordeiro D, Salas AI, Lodhia Z, Correia C, Isidro J, et al. Chlamydia trachomatis: When the virulence-associated genome backbone imports a prevalence-associated major antigen signature. Microb Genomics. 2019; 5(11). https://doi.org/10.1099/mgen.0.000313 PMID: 31697227

30. Cole MJ, Field N, Pitt R, Amato-Gauci AJ, Begovac J, French PD, et al. Substantial underdiagnosis of lymphogranuloma venereum in men who have sex with men in Europe: preliminary findings from a multicentre surveillance pilot. Sex Transm Infect. 2020; 96(2):137–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31235527 https://doi.org/10.1136/sextrans-2019-053972

31. Peuchant O, Touati A, Sperandio C, Hénin N, Laurier-Nadalié C, Bébéar C, et al. Changing Pattern of Chlamydia trachomatis Strains in Lymphogranuloma Venereum Outbreak, France, 2010–2015. Emerg Infect Dis. 2016 Nov; 22(11):1945–7. Available from: http://wwwnc.cdc.gov/eid/article/22/11/16-0247_article.htm PMID: 27767927

32. Rodríguez-Domínguez M, Puerta T, Menéndez B, González-Alba JM, Rodríguez C, Hellín T, et al. Clinical and epidemiological characterization of a lymphogranuloma venereum outbreak in Madrid, Spain: co-circulation of two variants. Clin Microbiol Infect. 2014 Mar; 20(3):219–25. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1198743X14608612 PMID: 23730727

33. Brunelle BW, Sensabaugh GF. The ompA gene in Chlamydia trachomatis differs in phylogeny and rate of evolution from other regions of the genome. Infect Immun. 2006 Jan; 74(1):578–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16369014 https://doi.org/10.1128/IAI.74.1.578-585.2006

34. Kari L, Whitmire WM, Crane DD, Reveneau N, Carlson JH, Goheen MM, et al. Chlamydia trachomatis native major outer membrane protein induces partial protection in nonhuman primates: implication for a trachoma transmission-blocking vaccine. J Immunol. 2009 Jun 15; 182(12):8063–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19494332 https://doi.org/10.4049/jimmunol.0804375

35. Brunelle BW, Sensabaugh GF. Nucleotide and phylogenetic analyses of the Chlamydia trachomatis ompA gene indicates it is a hotspot for mutation. BMC Res Notes. 2012 Jan 20; 5:53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22264291 https://doi.org/10.1186/1756-0500-5-53

36. Ferreira R, Antelo M, Nunes A, Borges V, Damião V, Borrego MJ, et al. In silico scrutiny of genes revealing phylogenetic congruence with clinical prevalence or tropism properties of Chlamydia trachomatis strains. G3 (Bethesda). 2014 Nov 5; 5(1):9–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25378473 https://doi.org/10.1534/g3.114.015354

37. Karl L, Whitmire WM, Carlson JH, Crane DD, Reveneau N, Nelson DE, et al. Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis. 2008 Feb 1; 197(3):449–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18199030 https://doi.org/10.1086/525285

38. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 2007 Oct; 13 Suppl 3:1–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17716294 https://doi.org/10.1111/j.1469-0691.2007.01786.x

39. Jeffrey BM, Suchland RJ, Quinn KL, Davidson JR, Stamm WE, Rockey DD. Genome sequencing of recent clinical Chlamydia trachomatis strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun. 2010 Jun; 78(6):2544–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20308297 https://doi.org/10.1128/IAI.01324-09

40. Quint KD, Bom RJ, Quint WGV, Bruisten SM, van der Loeff MFS, Morré SA, et al. Anal infections with concomitant Chlamydia trachomatis genotypes among men who have sex with men in Amsterdam, the Netherlands. BMC Infect Dis. 2011 Mar 14; 11:63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21401939 https://doi.org/10.1186/1471-2334-11-63

41. Chen C-Y, Chi K-H, Alexander S, Martin IMC, Liu H, Ison CA, et al. The molecular diagnosis of lymphogranuloma venereum: evaluation of a real-time multiplex polymerase chain reaction test using rectal and urethral specimens. Sex Transm Dis. 2007 Jul; 34(7):451–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17075436 https://doi.org/10.1097/01.olq.0000245957.02939.ea

42. Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, et al. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. Genome Res. 2007 Jan; 17(1):50–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17090662 https://doi.org/10.1101/gr.5674706

43. Joseph SJ, Didelot X, Gandhi K, Dean D, Read TD. Interplay of recombination and selection in the genomes of Chlamydia trachomatis. Biol Direct. 2011 May 26; 6:28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21615910 https://doi.org/10.1186/1745-6150-6-28

44. Somboonna N, Wan R, Ojcius DM, Pettengill MA, Joseph SJ, Chang A, et al. Hypervirulent Chlamydia trachomatis Clinical Strain Is a Recombinant between Lymphogranuloma Venereum (L2) and D Lineages. Campbell LA, Maloy S, editors. MBio. 2011 May 3; 2(3). Available from: https://mbio.asm.org/lookup/doi/10.1128/mBio.00045-11

45. Macdonald N, Ison C, Martin I, Alexander S, Lowndes C, Simms I, et al. Initial results of enhanced surveillance for lymphogranuloma venereum (LGV) in England. Euro Surveill. 2005 Jan 27; 10(1): E050127.5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16556955 https://doi.org/10.2807/esw.10.04.02629-en

APPENDIX 5 STBRL Protocol for LGV triplex PCR Assay

The LGV confirmation test used by NWLP and Viapath was originally developed by the Sexually Transmitted Bacteria Reference Laboratory (STBRL) at Public Health England (Morré *et al.*, 2005; Alexander, Martin and Ison, 2007; Chen *et al.*, 2007). The LGV test is a triplex PCR assay targeting a 36 bp deletion within the *pmpH* gene of LGV strains, in addition to an 88 bp region of the *C. trachomatis* cryptic plasmid, and the human RNase P gene as an internal process control. At NWLP, the protocol developed by STBRL is followed exactly; and at Viapath, the STBRL LGV assay has been adapted. Adaptations to the assay at Viapath are to the primers and probes, and these adaptations are highlighted in the relevant section. Sections of the STBRL triplex assay protocol have been reproduced with permission from the STBRL at Public Health England.

3.0 PERSONNEL

Procedure must only be carried out by suitably trained & qualified scientists. This includes technical and medical validation procedures

4.0 PRECAUTIONS

4.1 Specimens must be treated as potentially infectious substances and be handled accordingly.

4.2 Master Mix should be stored in light sensitive tubes – to ensure that the probe does not degrade (as the labelled probe is light sensitive).

4.3 Refer to G-6744 Minimising the risk of DNA contamination in Nucleic Acid Amplification Assays and monitoring for Environmental Contamination

5.0 METHOD

If a Master Mix has been made and aliquoted, proceed to 5.3; otherwise carry out procedure as outlined in 5.2.

5.1 **Equipment**

CT Forward primer (10 µM)
CT Reverse primer (10 µM)
CT Probe2 *HEX* (10 µM)
LGV forward primer (10 µM)
LGV reverse primer (10 µM)
LGV probe (10 µM)
RA Forward primer (10µM)
RA Reverse primer (10µM)
RA Probe (25µM)
$MgCl_2$ (25mM) *(Applied Biosystems).*
10 x PCR buffer *(Applied Biosystems).*
dNTP mix 12.5mM with dUTPs *(Applied Biosystems).*
Molecular grade water.
Amplitaq Gold (5 U μl^{-1}) *(Applied Biosystems).*
UNG (1 U μl^{-1}) *(Applied Biosystems).*
0.1 *or* 0.2 ml PCR tubes.
2ml Light sensitive tubes.
Pipettes capable of pipetting volumes of 1µl to 1000ul with filter tips.
PCR cabinets.
Rotorgene 3000, 6000 or *Rotorgene*-Q
Vortex.

5.2 **Preparation of *C. trachomatis* Specific Real-time Master Mix**

5.2.1 For each new batch of Master Mix allocate a batch number, this is obtained from form GW-0006

5.2.2 Defrost components and prepare PCR master mix according to table one (below). Those components highlighted in grey are not added at this stage.

Table One: *PCR Master Mix Composition*

Component	Volume required for one 25 µl reaction	Volume required for 10 reactions	*Final concentrations per 25 µl reaction*
CT Forward primer (10µM solution)	0.75µl	7.5µl	*0.3 µM*
CT Reverse primer (10µM Solution)	1µl	10µl	*0.4 µM*
CT probe-2 HEX (10µM solution)	0.75µl	7.5µl	*0.3 µM*
LGV Forward primer (10µM solution)	0.5µl	5µl	*0.2 µM*
LGV Reverse primer (10µM solution)	0.5µl	5µl	*0.2 µM*
LGV Probe (10µM solution)	0.625µl	6.25µl	*0.25 µM*
RA Forward Primer (10µm Solution)	0.2µl	2µl	*0.08 µM*
RA Reverse Primer (10µm Solution)	0.2µl	2µl	*0.08 µM*
RA Probe (25µm Solution)	0.2µl	2µl	*0.2 µM*
UNG (1U μl^{-1})	0.25µl	2.5µl	*0.25U*
AmpliTaq Gold (5U μl^{-1})	1µl	10µl	*5U*
$MgCl_2$ (25mM)	4µl	40µl	*4 mM*
10 x PCR buffer	2.5µl	25µl	*1x*
dNTP mix (12.5mM) with dUTPs	0.8µl	8µl	*0.4 mM*
Water	1.725µl	17.25µl	
Template	10µl	100µl	

5.2.3 **Vortex the master mix before aliquoting**

5.2.4 Aliquot master mix into light sensitive tubes, in aliquots sufficient for 40 reactions – (13.75µl x 40 +10µl extra = **560µl**)

5.2.5 Record all batch numbers for components (table one) on worksheet BRDW0042

5.2.6 Prepare labels for all aliquots with:

- Batch Number
- Reaction Name
- Expiry date

5.3 PCR Set Up

NOTE: Please use QIAgility instrument for all PCRs set-up where there are more than 15 clinical samples.

5.3.1 Calculate the quantity of master mix (13.75µl of master mix per sample), *Taq* (1µl per sample) and UNG (0.25µl) – depending on the number of samples and controls being tested. Also ensure extra reactions are made to compensate for minor fluctuations in pipetting volumes.

5.3.2 Defrost master mix or use directly if freshly prepared (5.2)

5.3.3 Vortex master mix and the enzymes - *Taq* and UNG.

5.3.4 Pipette the required volume of master mix into a sterile 1.5 ml starstedt tube, add the required volume of both *Taq* and UNG.

5.3.5 Briefly centrifuge the master mix (5.3.4) and aliquot 15 µl of the reaction mix into either 0.1 ml or 0.2 ml tubes.
Note: if 0.2 ml tubes are used ensure that they are flat capped tubes only; domed lid tubes are not compatible with the *Rotorgene*.

5.3.6 Record batch number of master mix, *Taq* and UNG on form BRDW-0003.

5.3.7 Add 10 µl of DNA lysate to master mix and gently pipette up and down three times.

5.3.8 Add 10 µl of;

- extraction negative controls,
- negative/non-template (water) control,
- *C.trachomatis LGV* positive control at a 10^{-1} dilution**.
- *C.trachomatis LGV* positive control at a 10^{-2} dilution**.

5.3.9 Place lids on tubes and incubate at room temperature in the dark for ten minutes for UNG activity to occur.

5.3.10 Place reaction tubes in to the correct Rotorgene rotor (36 well rotor if 0.2ml tubes are used: 72 well rotor if 0.1ml tubes are used). Ensure that any unused positions on the rotor are filled with empty tubes to ensure the rotor is balanced.

5.3.11 Use the quick start menu and open the folder 'SRS Templates with channels ' and open the ***CT-LGV multiplex*** programme and follow the online instructions (see G6771 for specific details).

5.4 Data Analysis

Following the completion of the real-time run the data can be examined

5.4.1 Click analysis on the tool-bar
Click on the **HEX** channel (*Rotorgene 3000*) or the **Yellow** (CT) channel (*Rotorgene 6000*)
Click Linear scale
Ensure that the Dynamic Tube button is selected.
Ensure that the 'Ignore first' is Set on 0

5.4.2 In the open **HEX** window (*Rotorgene 3000*) or the **Yellow** (CT) channel (*Rotorgene 6000*) use the manual threshold icon on the toolbar to set the base-line. *As a guide the threshold should be set at 0.1 (if any deviation from this guideline is required ensure that this is marked for the attention of the technical validator on BRDW0003 and give a reason for the deviation).*

5.4.3 Any specimens which generate a **HEX** fluorescence which exceeds the base-line should be recorded as positive for ***C. trachomatis.***
Note: Reports can be printed out using the report option.

5.4.4 Click analysis on the tool-bar & click on the **FAM** channel (*Rotorgene 3000*) or the **Green** (LGV) channel (*Rotorgene 6000*).
Click Linear scale
Ensure that the Dynamic Tube button is selected.
Ensure that the 'Ignore first' is Set on 0.

5.4.5 In the open **FAM** window (*Rotorgene 3000*) or the **Green** (LGV) channel (*Rotorgene 6000*) use the manual threshold icon on the toolbar to set the base-line. *As a guide the threshold should be set at 0.1 (if any deviation from this guideline is required ensure that this is marked for the attention of the technical validator on BRDW0003 and give a reason for the deviation).*

5.4.6 Any specimens which generate a **FAM** fluorescence which exceeds the base-line should be recorded as positive for **LGV**
Note: Reports can be printed out using the report option.

5.4.7 Click analysis on the tool-bar & click on the **Cy5** channel (*Rotorgene 3000*) or the **Red** (RNase-P) channel (*Rotorgene 6000*).
Click Linear scale
Ensure that the Dynamic Tube button is selected.
Ensure that the 'Ignore first' is Set on 0.

5.4.8 In the open **Cy5** channel (*Rotorgene 3000*) or the **Red** (RNAse-P) channel (*Rotorgene 6000*) (this channel detects the amplification of the internal control).

5.4.9 Any specimens which generate a **Cy5** fluorescence which exceeds the base-line have a positive amplification of the **RNase P gene.**

5.6 (Table 2) **Summary of results** (also see flow chart in the appendix)

Result in HEX/Yellow Channel *(C.trachomatis)*	**Result in FAM/Green Channel** *(LGV)*	**Result in CY5/Red Channel** *(RNase P)*	**Action**	**Result** (after any repeats)
Positive [Confirmation of Chlamydia only -for **Medico Legal** cases]	Positive or Negative	Positive	Perform Artus PCR G-6765	**Positive**
Negative [Confirmation of Chlamydia only -for **Medico Legal** cases]	Positive or Negative	Negative	Perform Artus PCR G-6765	**Inhibited**
Positive	Positive	Positive	Re-extract (if sufficient) & repeat PCR. - *for confirmation of LGV*	**CT & LGV Positive**
Positive	Negative	Positive	Report as: Chlamydia positive LGV negative	**CT Positive LGV Negative**
Negative	Negative	Positive	Re-extract (if sufficient) & repeat PCR	**Negative**
Negative	Positive	Positive	Re-extract (if sufficient) & repeat PCR. if same result perform Artus PCR (G-6765)	**LGV positive**
Positive	Positive	Negative	Re-extract (if sufficient) & repeat PCR - *for confirmation of LGV*	**LGV positive**
Positive	Negative	Negative	Re-extract (if sufficient) & repeat PCR.	**possibly LGV inhibited**
Negative	Positive	Negative	Re-extract (if sufficient) & repeat PCR.	**possibly CT inhibited**
Negative	Negative	Negative	Re-extract (if sufficient) & repeat PCR.	**Inhibited**
C_t value of ≥ 40		Positive	Re-extract (if sufficient) & repeat PCR. if same report as equivocal **Write Ct value in the MOLIS internal comments**	**Equivocal**
C_t value of ≥ 40		Negative	Re-extract (if sufficient) & repeat PCR.	**probable inhibition**

Any equivocal results should be repeated.

Primers and Probes used at NWLP (and STBRL):

Table 4:
Primer and Probe sequences.

primer / probe	Sequence 5' – 3'	Target
CTF-008	**GGATTGACTCCGACAACGTATTC**	*C.trachomatis* cryptic plasmid
CTR-009	**ATCATTGCCATTAGAAAGGGCATT**	
CT-Probe 2 HEX	*HEX*-**TTACGTGTAGGCGGTTTAGAAAGCGG**-*bhq1*	
LGV-F	**CTG TGC CAA CCT CAT CAT CAA**	Polymorphic membrane protein *H*
LGV-R	**AGA CCC TTT CCG AGC ATC ACT**	
LGV-Probe	*6FAM*-**CCT GCT CCA ACA GT**-*mgb*	
RA forward	**AGATTTGGACCTGCGAGCG**	Human RNase P gene
RA reverse	**GAGCGGCTGTCTCCACAAGT**	
RA probe	*Cy5*-**TTCTGACCTGAAGGCTCTGCGCG**- *bhq3*	

Note: primers and Probes ordered from Applied-Biosystems, Sigma or suitable alternative.

Table 5:
Reaction Conditions.

Temp (ºC)	Time	*Acquisition*	No. Cycles
50ºC	**2 mins**		**1**
95ºC	**10 mins**		**1**
95ºC	**15 Sec**		**40**
60ºC	**60 sec**	**FAM, HEX, Cy5**	

Primers and Probes Used at Viapath:

Name	Sequence	Additional information	Supplier
CtF	GGATTGACTCCGACAACGTATTC		Eurogentec
CtR	ATCATTGCCATTAGAAAGGGCATT		Eurogentec
CtProbe	Cy5-TTACGTGTAGGCGGTTTAGAAAGCGG-bhq2	Hex label used at Colindale	Eurogentec
LGVF	CTGTGCCAACCTCATCATCAA		Eurogentec
LGVR	AGACCCTTTCCGAGCATCACT		Eurogentec
LGVProbe	6FAM-CCTGCTCCAACAGT-mgb and 6FAM-CTTGCTCCAACAGT-mgb	Minor Groove Binder	Life Technologies
RaF	AGATTTGGACCTGCGAGCG		Eurogentec
RaR	GAGCGGCTGTCTCCACAAGT		Eurogentec
RaProbe	Hex-TTCTGACCTGAAGGCTCTGCGCG-bhq1	Cy5 Label used at Colindale	Eurogentec

APPENDIX 6 Specimen Type, Collection Date & GUM Clinic of Origin of DNA Extracts from NWLP and Viapath

DNA Extracts from NWLP ***EQA = External Quality Assurance**

Sample Number	Specimen Type	Date Collected from patient	GUM Clinic
001	Rectal Swab	05/04/2019	56 Dean Street
002	Rectal Swab	05/04/2019	56 Dean Street
003	Rectal Swab	07/11/2018	56 Dean Street
004	Rectal Swab	07/11/2018	56 Dean Street
005*		07/11/2018	EQA sample
006	Rectal Swab	07/11/2018	56 Dean Street
007	Rectal Swab	06/11/2018	56 Dean Street
008	Rectal Swab	06/11/2018	56 Dean Street
009	Rectal Swab	11/03/2019	56 Dean Street
010	Rectal Swab	11/03/2019	56 Dean Street
011	Rectal Swab	11/03/2019	56 Dean Street
012	Rectal Swab	11/03/2019	56 Dean Street
013	Rectal Swab	03/06/2019	56 Dean Street
014	Rectal Swab	31/05/2019	10 Hammersmith Broadway
015	Rectal Swab	03/06/2019	56 Dean Street
016	Rectal Swab	03/06/2019	56 Dean Street
017	Rectal Swab	03/06/2019	56 Dean Street
018	Rectal Swab	03/06/2019	56 Dean Street
019	Rectal Swab	03/06/2019	56 Dean Street
020	Rectal Swab	03/06/2019	56 Dean Street
021	Rectal Swab	03/06/2019	56 Dean Street
022	Rectal Swab	04/06/2019	56 Dean Street
023	Rectal Swab	15/02/2019	56 Dean Street
024	Rectal Swab	15/02/2019	56 Dean Street
025	Rectal Swab	15/02/2019	56 Dean Street
026	Rectal Swab	15/02/2019	56 Dean Street
027	Rectal Swab	24/01/2019	56 Dean Street
028	Rectal Swab	24/01/2019	56 Dean Street
029	Rectal Swab	24/01/2019	56 Dean Street
030	Rectal Swab	24/01/2019	Jefferiss Wing
031	Rectal Swab	28/01/2019	56 Dean Street
032	Rectal Swab	28/01/2019	56 Dean Street
033	Rectal Swab	28/01/2019	56 Dean Street
034	Rectal Swab	01/02/2019	Jefferiss Wing

035	Rectal Swab	05/03/2019	56 Dean Street
036	Rectal Swab	23/05/2019	10 Hammersmith Broadway
037	Rectal Swab	21/05/2019	56 Dean Street
038	Rectal Swab	18/05/2019	56 Dean Street
039	Rectal Swab	20/05/2019	56 Dean Street
040	Rectal Swab	06/07/2018	56 Dean Street
041	Rectal Swab	04/07/2018	Jefferiss Wing
042	Rectal Swab	28/06/2018	John Hunter Clinic
043	Rectal Swab	29/06/2018	56 Dean Street
044	Rectal Swab	28/06/2018	56 Dean Street
045	Rectal Swab	26/06/2018	56 Dean Street
046	Rectal Swab	22/06/2018	56 Dean Street
047	Rectal Swab	20/03/2019	56 Dean Street
048	Rectal Swab	20/03/2019	56 Dean Street
049	Rectal Swab	12/03/2019	Jefferiss Wing
050	Rectal Swab	18/03/2019	56 Dean Street
051	Rectal Swab	18/03/2019	John Hunter Clinic
052	Rectal Swab	19/03/2019	56 Dean Street
053	Rectal Swab	26/02/2019	56 Dean Street
054	Rectal Swab	26/02/2019	56 Dean Street
055	Rectal Swab	26/02/2019	56 Dean Street
056	Rectal Swab	27/02/2019	56 Dean Street
057	Rectal Swab	28/02/2019	56 Dean Street
058	Rectal Swab	06/02/2019	Jefferiss Wing
059	Rectal Swab	12/04/2019	10 Hammersmith Broadway
060	Rectal Swab	12/04/2019	10 Hammersmith Broadway
061	Rectal Swab	16/01/2019	56 Dean Street
062	Rectal Swab	15/01/2019	56 Dean Street
063	Rectal Swab	14/01/2019	Jefferiss Wing
064	Rectal Swab	15/01/2019	56 Dean Street
065	Rectal Swab	15/01/2019	56 Dean Street
066	Rectal Swab	16/01/2019	56 Dean Street
067	Rectal Swab	16/01/2019	56 Dean Street
068	Rectal Swab	15/01/2019	56 Dean Street
069	Rectal Swab	15/01/2019	56 Dean Street
070	Rectal Swab	20/02/2019	56 Dean Street
071	Rectal Swab	20/02/2019	56 Dean Street
072	Rectal Swab	18/02/2019	56 Dean Street
073	Rectal Swab	14/02/2019	John Hunter Clinic
074	Rectal Swab	19/02/2019	56 Dean Street
075	Rectal Swab	19/02/2019	56 Dean Street

076	Rectal Swab	04/06/2019	56 Dean Street
077	Rectal Swab	06/06/2019	56 Dean Street
078	Rectal Swab	21/05/2019	56 Dean Street
079	Rectal Swab	17/05/2019	56 Dean Street
080	Rectal Swab	16/05/2019	56 Dean Street
081	Rectal Swab	21/05/2019	56 Dean Street
082	Rectal Swab	16/05/2019	56 Dean Street
083	Rectal Swab	14/05/2019	56 Dean Street
084	Rectal Swab	17/05/2019	56 Dean Street
085	Rectal Swab	14/03/2019	10 Hammersmith Broadway
086	Rectal Swab	12/03/2019	56 Dean Street
087	Rectal Swab	12/03/2019	56 Dean Street
088	Rectal Swab	12/03/2019	56 Dean Street
089	Rectal Swab	12/03/2019	56 Dean Street
090	Rectal Swab	13/03/2019	Not known
091	Rectal Swab	13/03/2019	56 Dean Street
092	Rectal Swab	23/04/2019	Jefferiss Wing
093	Rectal Swab	29/04/2019	56 Dean Street
094	Rectal Swab	26/04/2019	John Hunter Clinic
095	Rectal Swab	29/04/2019	56 Dean Street
096	Rectal Swab	11/05/2019	56 Dean Street
097	Rectal Swab	11/05/2019	56 Dean Street
098	Rectal Swab	04/05/2019	56 Dean Street
099	Rectal Swab	29/05/2019	56 Dean Street
100	Rectal Swab	20/05/2019	56 Dean Street
101	Rectal Swab	24/05/2019	56 Dean Street
102	Rectal Swab	02/04/2019	56 Dean Street
103	Rectal Swab	01/04/2019	56 Dean Street
104	Rectal Swab	30/03/2019	56 Dean Street
105	Rectal Swab	29/03/2019	56 Dean Street
106	Rectal Swab	06/06/2019	56 Dean Street
107	Rectal Swab	06/06/2019	56 Dean Street
108	Rectal Swab	07/06/2019	56 Dean Street
109	Rectal Swab	05/06/2019	56 Dean Street
110*		11/06/2019	EQA Sample
111	Rectal Swab	20/10/2018	56 Dean Street
112	Rectal Swab	20/10/2018	56 Dean Street
113	Rectal Swab	12/10/2018	10 Hammersmith Broadway
114	Rectal Swab	24/09/2018	56 Dean Street
115	Rectal Swab	18/10/2018	56 Dean Street
116	Rectal Swab	18/10/2018	56 Dean Street

117	Rectal Swab	18/10/2018	56 Dean Street
118	Rectal Swab	18/10/2018	56 Dean Street
119	Rectal Swab	24/11/2018	56 Dean Street
120	Rectal Swab	23/11/2018	56 Dean Street
121	Rectal Swab	21/11/2018	John Hunter Clinic
122	Rectal Swab	26/11/2018	56 Dean Street
123	Rectal Swab	11/08/2018	56 Dean Street
124	Rectal Swab	13/08/2018	56 Dean Street
125	Rectal Swab	11/08/2018	56 Dean Street
126	Rectal Swab	13/08/2018	56 Dean Street
127	Rectal Swab	10/08/2018	56 Dean Street
128	Rectal Swab	09/08/2018	56 Dean Street
129	Rectal Swab	10/08/2018	56 Dean Street
130	Rectal Swab	09/08/2018	56 Dean Street
131	Rectal Swab	06/08/2018	John Hunter Clinic
132	Rectal Swab	07/08/2018	56 Dean Street
133	Rectal Swab	14/03/2019	10 Hammersmith Broadway
134*		03/08/2018	EQA Sample
135	Rectal Swab	03/08/2018	56 Dean Street
136	Rectal Swab	20/03/2019	56 Dean Street
137	Rectal Swab	21/03/2019	56 Dean Street
138	Rectal Swab	31/01/2019	56 Dean Street
139	Rectal Swab	31/01/2019	56 Dean Street
140	Rectal Swab	31/01/2019	56 Dean Street
141	Rectal Swab	31/01/2019	56 Dean Street
142	Rectal Swab	16/05/2019	56 Dean Street
143	Rectal Swab	16/05/2019	56 Dean Street
144	Rectal Swab	15/05/2019	10 Hammersmith Broadway
145	Rectal Swab	13/05/2019	56 Dean Street
146	Rectal Swab	23/05/2019	56 Dean Street
147	Rectal Swab	23/05/2019	56 Dean Street
148	Rectal Swab	23/05/2019	56 Dean Street
149	Rectal Swab	12/04/2019	56 Dean Street
150	Rectal Swab	16/04/2019	10 Hammersmith Broadway
151	Rectal Swab	12/04/2019	56 Dean Street
152	Rectal Swab	15/04/2019	56 Dean Street
153	Rectal Swab	12/04/2019	56 Dean Street
154	Rectal Swab	03/01/2019	56 Dean Street
155	Rectal Swab	03/01/2019	56 Dean Street
156	Rectal Swab	03/01/2019	56 Dean Street
157	Rectal Swab	03/01/2019	56 Dean Street

158	Rectal Swab	12/07/2018	56 Dean Street
159	Rectal Swab	13/07/2018	56 Dean Street
160	Rectal Swab	10/07/2018	10 Hammersmith Broadway
161	Rectal Swab	28/12/2018	56 Dean Street
162	Rectal Swab	28/12/2018	56 Dean Street
163	Rectal Swab	28/12/2018	56 Dean Street
164	Rectal Swab	28/12/2018	56 Dean Street
165	Rectal Swab	28/12/2018	56 Dean Street
166	Rectal Swab	28/12/2018	56 Dean Street
167	Rectal Swab	31/12/2018	56 Dean Street
168	Rectal Swab	31/12/2018	56 Dean Street
169	Rectal Swab	28/12/2018	56 Dean Street
170	Rectal Swab	24/08/2018	56 Dean Street
171	Rectal Swab	24/08/2018	56 Dean Street
172	Rectal Swab	22/08/2018	56 Dean Street
173	Rectal Swab	22/08/2018	56 Dean Street
174	Rectal Swab	21/08/2018	10 Hammersmith Broadway
175	Rectal Swab	21/08/2018	56 Dean Street
176	Rectal Swab	06/06/2019	56 Dean Street
177	Rectal Swab	06/06/2019	56 Dean Street
178	Rectal Swab	07/06/2019	56 Dean Street
179	Rectal Swab	05/06/2019	56 Dean Street
180*		11/06/2019	EQA Sample

DNA Extracts from Viapath ***GUM clinic information not obtained.**

Sample Number	Specimen Type	Date Collected from patient
001	3in1	28/08/2018
002	Rectal Swab	26/08/2018
003	3in1	16/08/2018
004	Urine	09/08/2018
005	3in1	14/08/2018
006	Rectal Swab	25/07/2018
007	Rectal Swab	30/07/2018
008	3in1	23/07/2018
009	3in1	19/07/2018
010	Rectal Swab	19/07/2018
011	Rectal Swab	23/07/2018
012	Throat Swab	23/07/2018
013	Rectal Swab	17/07/2018

014	3in1	12/07/2018
015	Rectal Swab	13/07/2018
016	3in1	11/07/2018
017	Rectal Swab	07/07/2018
018	Rectal Swab	02/07/2018
019	Rectal Swab	21/06/2018
020	Rectal Swab	22/06/2018
021	Urine	22/06/2018
022	3in1	14/06/2018
023	Rectal Swab	15/06/2018
024	Rectal Swab	07/06/2018
025	3in1	07/06/2018
026	Rectal Swab	08/06/2018
027	3in1	31/05/2018
028	3in1	24/05/2018
029	3in1	25/05/2018
030	Rectal Swab	26/05/2018
031	Rectal Swab	17/05/2018
032	Rectal Swab	16/05/2018
033	Rectal Swab	18/05/2018
034	3in1	08/05/2018
035	Rectal Swab	25/04/2018
036	3in1	29/04/2018
037	3in1	19/04/2018
038	3in1	21/04/2018
039	Rectal Swab	15/06/2018
040	Rectal Swab	15/03/2018
041	3in1	19/03/2018
042	3in1	02/03/2018
043	3in1	28/03/2018
044	3in1	23/02/2018
045	3in1	23/02/2018
046	Rectal Swab	24/02/2018
047	3in1	27/02/2018
048	Rectal Swab	15/05/2018
049	Rectal Swab	16/02/2018
050	3in1	17/02/2018

APPENDIX 7 MLVA-*ompA* Genotypes Detected in London

Clinical specimens were collected from patients by clinicians or nurses at each GUM clinic and sent to Viapath or North West London Pathology (NWLP) for *C. trachomatis* testing. Specimens collected from Burrell Street Sexual Health Centre, Streatham Hill Clinic, Walworth Road Clinic, and Harrison Wing were tested for *C. trachomatis* at Viapath. Specimens collected from 56 Dean Street, John Hunter Clinic, 10 Hammersmith Broadway, and Jefferiss Wing were tested for *C. trachomatis* at NWLP. DNA extracts were numbered from 001 onwards and anonymised by laboratory personnel at NWLP and Viapath. VNTR type codes for each of the three VNTR loci (CT1335, CT1299, and CT1291) and the *ompA* genotype were determined for each DNA extract. Black boxes indicate that a genotype could not be assigned at the given locus. Asterisks (*) indicate NWLP extracts that were EQA samples (i.e. external quality assurance samples; not patient samples).

NWLP					VIAPATH				
Extract	**CT1335**	**CT1299**	**CT1291**	***ompA***	**Extract**	**CT1335**	**CT1299**	**CT1291**	***ompA***
001	1	9	2b	L2	001	1			
002	1	9	2b	L2b	002	1	9	2b	L2
003	1		2b	L2b	003				
004	1	9	2b	L2	004	1	9	2b	
005*	1	9	2b	L2	005	1	9	2b	L2b
006	1			L2bV6	006	1	9	2b	
007		9	3b		007	1	9	2b	L2
008	1	9	2b	L2b	008	1	9		L2
009	1	9	2b	L2	009				
010	1	9	2b	L2b	010	1	9	2b	L2
011	1	9	2b	L2b	011	1	9	2b	L2
012	1	9	2b	L2bv5	012	1	9	2b	L2
013					013	1	9	2b	L2
014	1	9	2b	L2bV5	014	1	9	2b	L2
015	1	9	2b	L2	015				
016	1	9	3b	L2b	016	1	9	3b	
017	8	5		E	017	1	9		
018	1	9	2b	L2b/D	018				
019	1	9	2b	L2	019	1	9	2b	L2
020	3		3	G	020				
021	1	9	2b	L2b	021				
022		9	2b	L2	022				
023	1	9	3b	L2b	023	1	9	2b	L2b
024	1	9	2b	L2	024	1	9	2b	L2b
025	1	9	2b	L2b	025	1	9	3b	L2bV5

026				
027	1	9	2b	L2b
028				
029	1	9	2b	L2b
030				
031	1	9	2b	L2b
032	1	9	2b	L2b
033	1	9	2b	
034	1	9	2b	L2
035	1	9	2b	L2
036	5	9	2b	L1
037	1	9	2b	L2b
038	8	6		E
039	1	9	2b	L2
040	1	9	3b	L2bV5
041	1	9	3b	L2bV5
042	1	9	2b	L2b/D
043	1	9	2b	L2bV5
044	1	9	2b	L2
045				
046	1	9	2b	L2
047	1	9	3b	L2bV6
048	1	9	2b	L2
049	1	9	2b	L2b
050				
051	1	9	2b	L2b
052		9	2b	L2
053	1	9	2b	L2b
054				
055	1	9	2b	
056	1	9	2b	L2b
057	1	9		
058	1	9	2b	L2b
059	1	9	2b	L2
060	1	9	2b	L2b
061		9	2b	
062	1	9	3b	L2bV6
063	1	9	2b	L2b
064	1	9	2b	L2
065	3	9	4	
066	3	9	2b	

026	8	6		E
027	1	9	3b	L2bV5
028	1	9	2b	
029	1	9	2b	L2
030	1	9		
031	1	9	2b	L2b/D
032	1			
033		8		
034	1	9	3b	
035	1	9	2b	L2
036	1	9	2b	
037	1	9	2b	L2b
038	1	9	2b	
039	1	9	2b	L2
040	1	9	2b	
041	1	9	2b	
042	1	9	2b	L2
043	1	9	2b	L2
044				
045	1	9	2b	L2
046	1	9	2b	L2
047	1	9	3b	L2bV5
048	1	9	2b	L2h
049	1	9	2b	
050	1	9	2b	

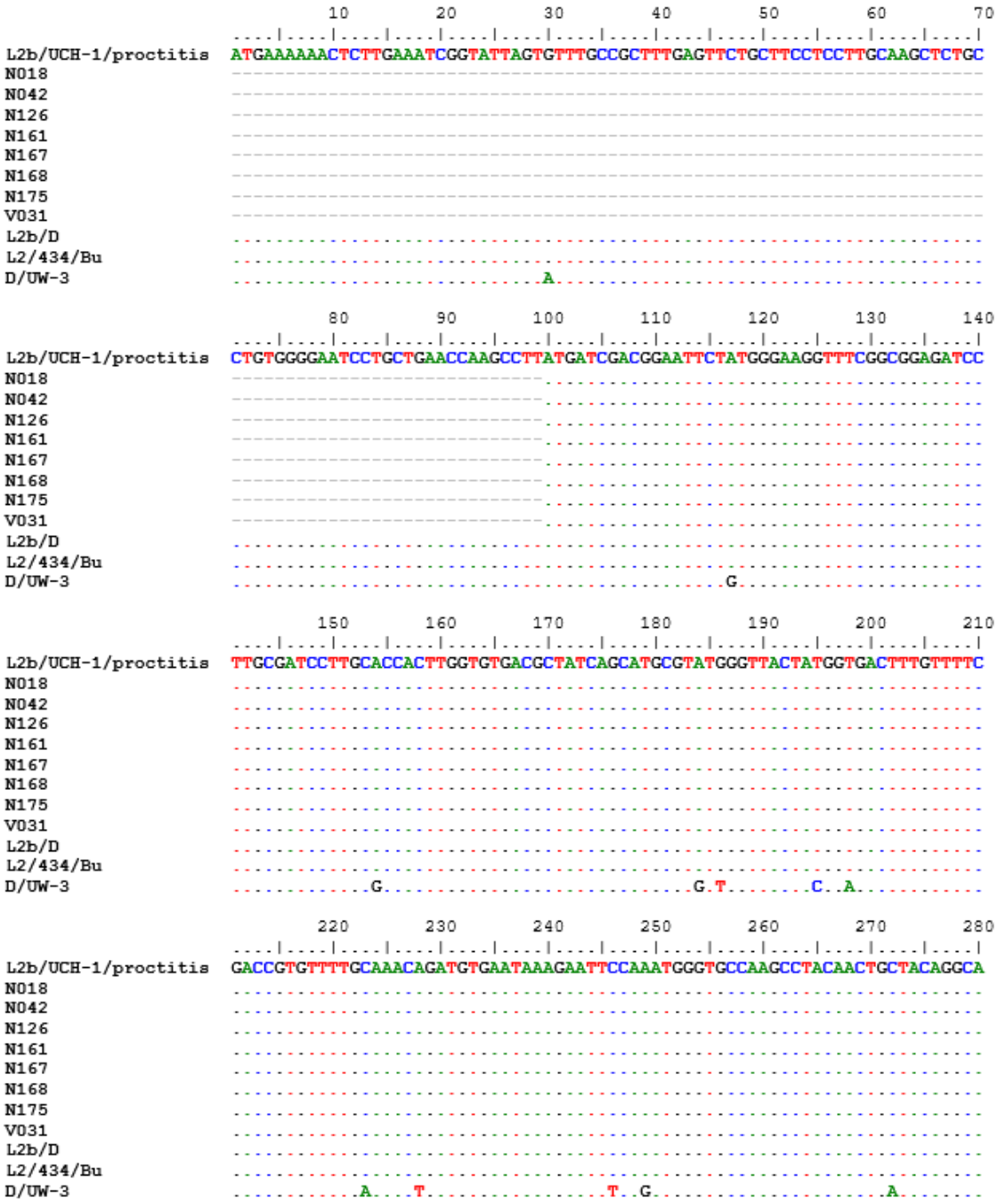
067	1	9	2b	L2
068	1	4a	2b	L2b
069	1	9	2b	L2
070	1	9	2b	L2b
071	1	9	2b	L2bV5
072	1	9	2b	L2b
073	1	9	2b	L2b
074	1	9	2b	L2
075	1	9	2b	L2b
076	1	9		
077	1	9	3b	
078	1	9	2b	L2b
079	1	9	2b	L2
080	1	9	2b	L2b
081	1	9	2b	L2
082	1	9	2b	L2b
083	1	9	2b	L2
084	1	9	2b	L2
085				
086	1	9	3b	L2bV6
087	1	9	2b	L2
088	1	9	2b	L2
089	1	9	2b	L2
090	1	9	2b	
091	1	9	2b	
092	1	9	2b	
093	1	9	2b	L2b
094	1	9	2b	L2b
095	1	9	2b	L2bV5
096	5	9	2b	L1
097	1	9	2b	L2bV5
098	1	9	2b	L2b
099	1	9	2b	L2
100	1	9	2b	L2b
101	1	9	3b	L2b
102	1	9	2b	L2
103	1	9	2b	L2b
104				
105	1	9	2b	
106	1	9	2b	L2
107	1	9	2b	L2

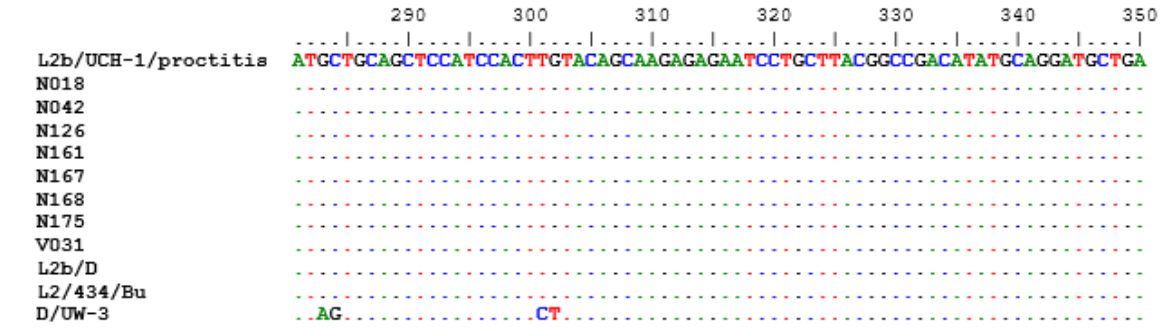
108	1	9	2b	L2
109				
110*	1	9	2b	L2
111	1	9	2b	L2
112	1	9	2b	L2
113				
114				
115	1	9	3b	L2bV5
116	1	9	2b	L2b
117	1	9	2b	L2
118	1	9	2b	L2
119	1	9	2b	L2b
120			2b	
121	1	9	2b	L2
122	1	9	2b	L2b
123	1	9	3b	L2bV5
124	1	9	2b	
125	3	9	3	G
126	1	9	2b	L2b/D
127				
128	1	9	2b	L2b
129	3	9	3	G
130	1	9	3b	L2bV5
131	1			
132				
133				
134*				
135	1	9	2b	L2b
136	1	9	2b	L2bV5
137	1	9	2b	L2
138	1	9	2b	L2
139	1	9	2b	L2
140	1	9	2b	L2b
141	1	9	3b	L2bV6
142	1	9	2b	L2b
143	1	9	2b	L2bV5
144	1	9	2b	L2
145	1	9	2b	L2
146	1	9	2b	L2
147	1	9	2b	L2
148	1	9	2b	L2bV5

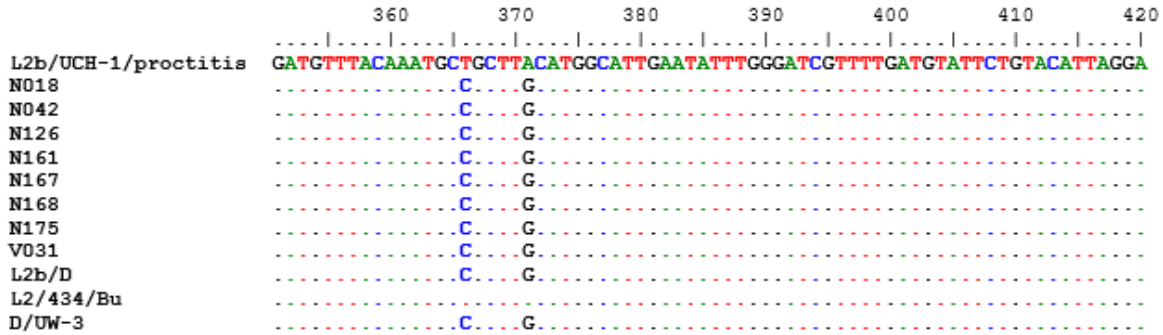
149	1	9	2b	L2
150	1	9	2b	L2
151	1	9	2b	L2
152	1	9	2b	L2
153	1	9	2b	L2
154	1	9	2b	L2
155	1	9	2b	L2
156				
157	1	9	2b	L2bV5
158	1	9	2b	L2
159	1	9	2b	L2
160	1	9	2b	L2
161	1	9	2b	L2b/D
162	1	9	2b	L2
163	1	9	2b	L2b
164	1	9	2b	L2
165	1	9	2b	L2
166	1	9	2b	L2b
167	1	9	2b	L2b/D
168	1	9	2b	L2b/D
169	1	9	2b	L2
170	1	9	2b	L2
171	1	9	2b	L2
172	3		3	G
173	1	9	2b	L2
174	5	9	2b	L1
175	1	9	2b	L2b/D
176	1	9	2b	L2
177	3		3	J
178	1	9	2b	L2
179	1	9	3b	L2bV6
180*	3	9	3	G

APPENDIX 8 Alignment of the *ompA* gene from DNA Extracts with a Hybrid L2b/D Profile

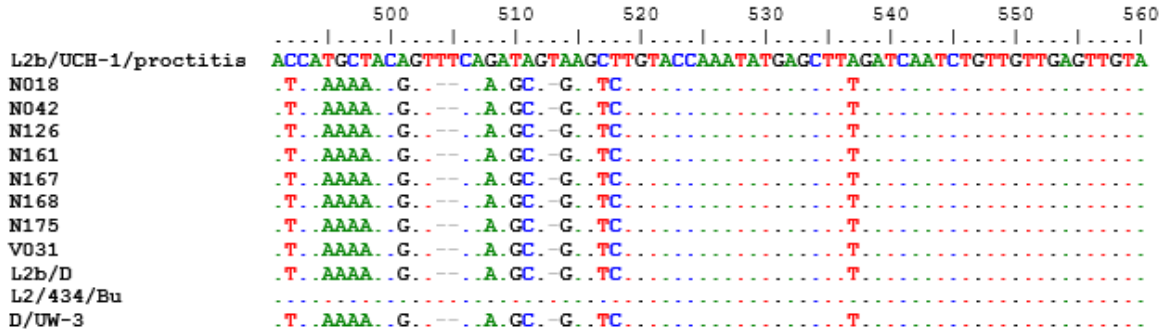
The *ompA* sequence of L2b/UCH-1/proctitis (Genbank accession no. AM884177.1) was used as a reference. Nucleotide numbers are given according to L2b/UCH-1/proctitis. L2b/D refers to the *ompA* sequence of the Portuguese L2b/D strain (Genbank Accession No. MN094864.1). The *ompA* sequences of L2/434/Bu (Genbank Accession No. AM884176.1) and D/UW-3 (Genbank Accession No. NC_000117.1) were included in the alignment. The ompA sequences of the 8 extracts assigned ompA genotype L2b/D in the genotyping study are included. Coloured dots indicate bases that matched L2b/UCH-1/proctitis.

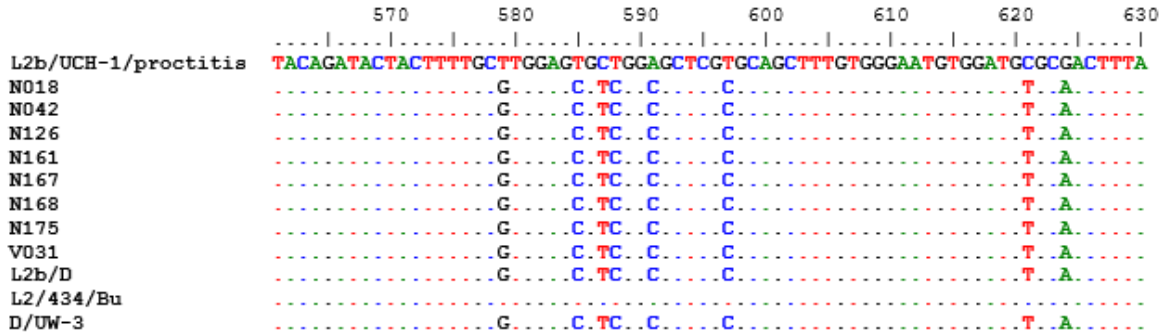


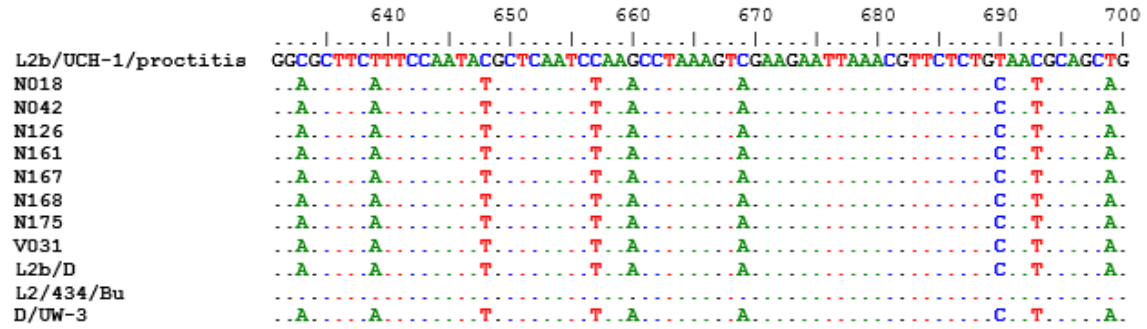

290 300 310 320 330 340 350
L2b/UCH-1/proctitis ATGCTGCAGCTCCATCCACTTGTACAGCAAGAGAGAATCCTGCTTACGGCCGACATATGCAGGATGCTGA
N018
N042
N126
N161
N167
N168
N175
V031
L2b/D
L2/434/Bu
D/UW-3 ..AG..........CT..


360 370 380 390 400 410 420
L2b/UCH-1/proctitis GATGTTTACAAATGCTGCTTACATGGCATTGAATATTTGGGATCGTTTTGATGTATTCTGTACATTAGGA
N018 .....C....G.....
N042 .....C....G.....
N126 .....C....G.....
N161 .....C....G.....
N167 .....C....G.....
N168 .....C....G.....
N175 .....C....G.....
V031 .....C....G.....
L2b/D .....C....G.....
L2/434/Bu
D/UW-3 .....C....G.....

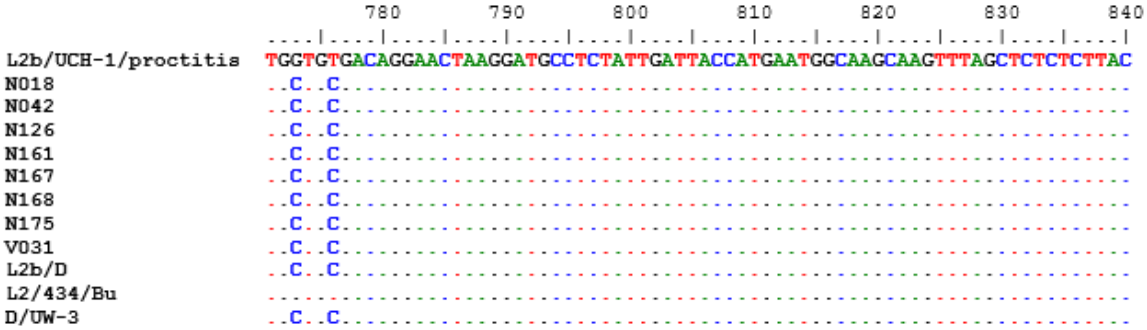

430 440 450 460 470 480 490
L2b/UCH-1/proctitis GCCACCAGTGGATATCTTAAAGGAAATTCAGCATCTTTCAACTTAGTTGGGTTATTCGGAGATAGTGAGA
N018 .....C..T..T.....T.....A..G..T.....A...A.
N042 .....C..T..T.....T.....A..G..T.....A...A.
N126 .....C..T..T.....T.....A..G..T.....A...A.
N161 .....C..T..T.....T.....A..G..T.....A...A.
N167 .....C..T..T.....T.....A..G..T.....A...A.
N168 .....C..T..T.....T.....A..G..T.....A...A.
N175 .....C..T..T.....T.....A..G..T.....A...A.
V031 .....C..T..T.....T.....A..G..T.....A...A.
L2b/D .....C..T..T.....T.....A..G..T.....A...A.
L2/434/Bu .....A.....
D/UW-3 .....C..T..T.....T.....A..G..T.....A...A.

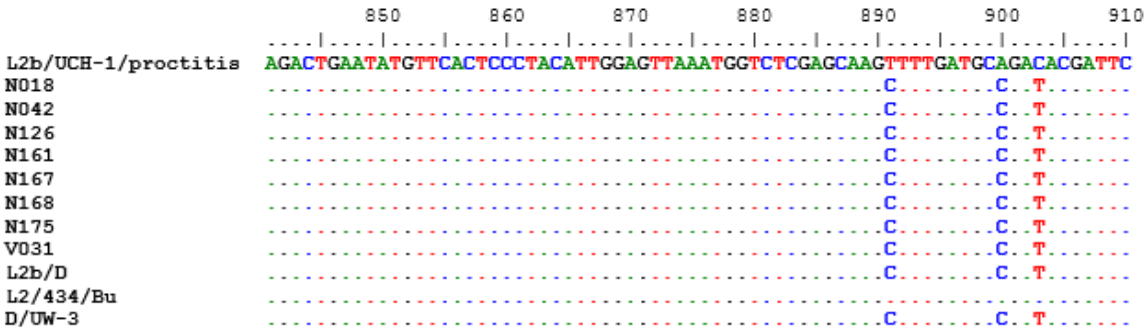

500 510 520 530 540 550 560
L2b/UCH-1/proctitis ACCATGCTACAGTTTCAGATAGTAAGCTTGTACCAAATATGAGCTTAGATCAATCTGTTGTTGAGTTGTA
N018 .T..AAAA..G..--..A.GC..G..TC.....T.....
N042 .T..AAAA..G..--..A.GC..G..TC.....T.....
N126 .T..AAAA..G..--..A.GC..G..TC.....T.....
N161 .T..AAAA..G..--..A.GC..G..TC.....T.....
N167 .T..AAAA..G..--..A.GC..G..TC.....T.....
N168 .T..AAAA..G..--..A.GC..G..TC.....T.....
N175 .T..AAAA..G..--..A.GC..G..TC.....T.....
V031 .T..AAAA..G..--..A.GC..G..TC.....T.....
L2b/D .T..AAAA..G..--..A.GC..G..TC.....T.....
L2/434/Bu
D/UW-3 .T..AAAA..G..--..A.GC..G..TC.....T.....

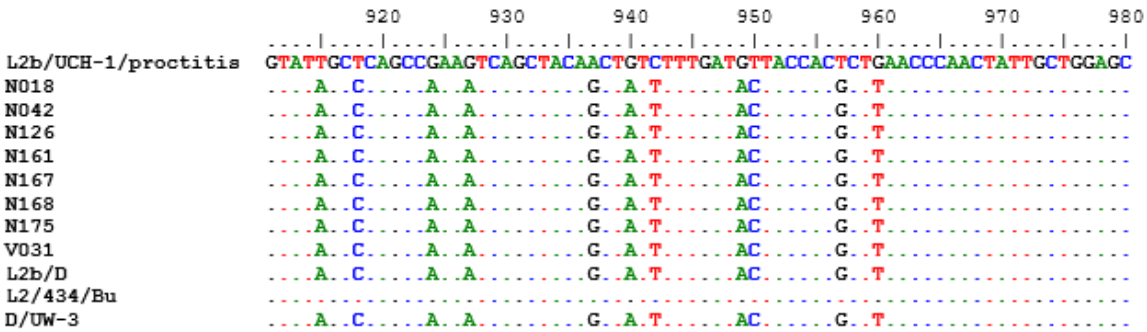

570 580 590 600 610 620 630
L2b/UCH-1/proctitis TACAGATACTACTTTTGCTTGGAGTGCTGGAGCTCGTGCAGCTTTGTGGGAATGTGGATGCGCGACTTTA
N018 .....G.....C.TC..C.....C.....T..A.....
N042 .....G.....C.TC..C.....C.....T..A.....
N126 .....G.....C.TC..C.....C.....T..A.....
N161 .....G.....C.TC..C.....C.....T..A.....
N167 .....G.....C.TC..C.....C.....T..A.....
N168 .....G.....C.TC..C.....C.....T..A.....
N175 .....G.....C.TC..C.....C.....T..A.....
V031 .....G.....C.TC..C.....C.....T..A.....
L2b/D .....G.....C.TC..C.....C.....T..A.....
L2/434/Bu
D/UW-3 .....G.....C.TC..C.....C.....T..A.....

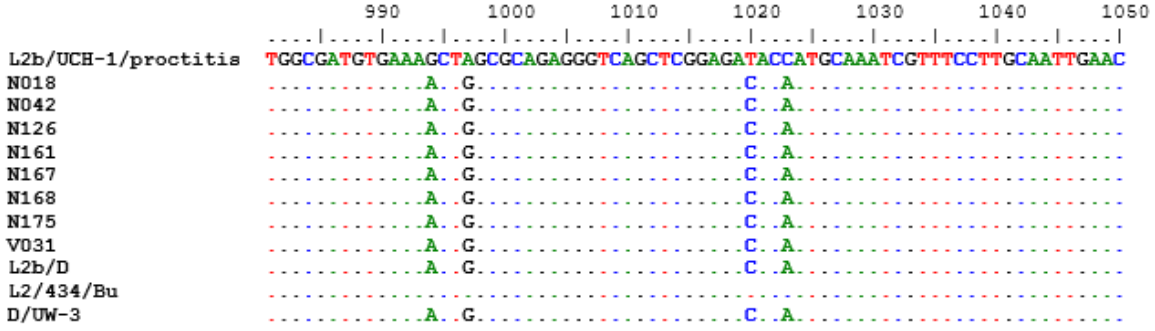

640 650 660 670 680 690 700
L2b/UCH-1/proctitis GGCGCTTCTTTCCAATACGCTCAATCCAAGCCTAAAGTCGAAGAATTAAACGTTCTCTGTAACGCAGCTG
N018 ..A....A.....T.....T..A.....A.....C..T.....A.
N042 ..A....A.....T.....T..A.....A.....C..T.....A.
N126 ..A....A.....T.....T..A.....A.....C..T.....A.
N161 ..A....A.....T.....T..A.....A.....C..T.....A.
N167 ..A....A.....T.....T..A.....A.....C..T.....A.
N168 ..A....A.....T.....T..A.....A.....C..T.....A.
N175 ..A....A.....T.....T..A.....A.....C..T.....A.
V031 ..A....A.....T.....T..A.....A.....C..T.....A.
L2b/D ..A....A.....T.....T..A.....A.....C..T.....A.
L2/434/Bu .....
D/UW-3 ..A....A.....T.....T..A.....A.....C..T.....A.

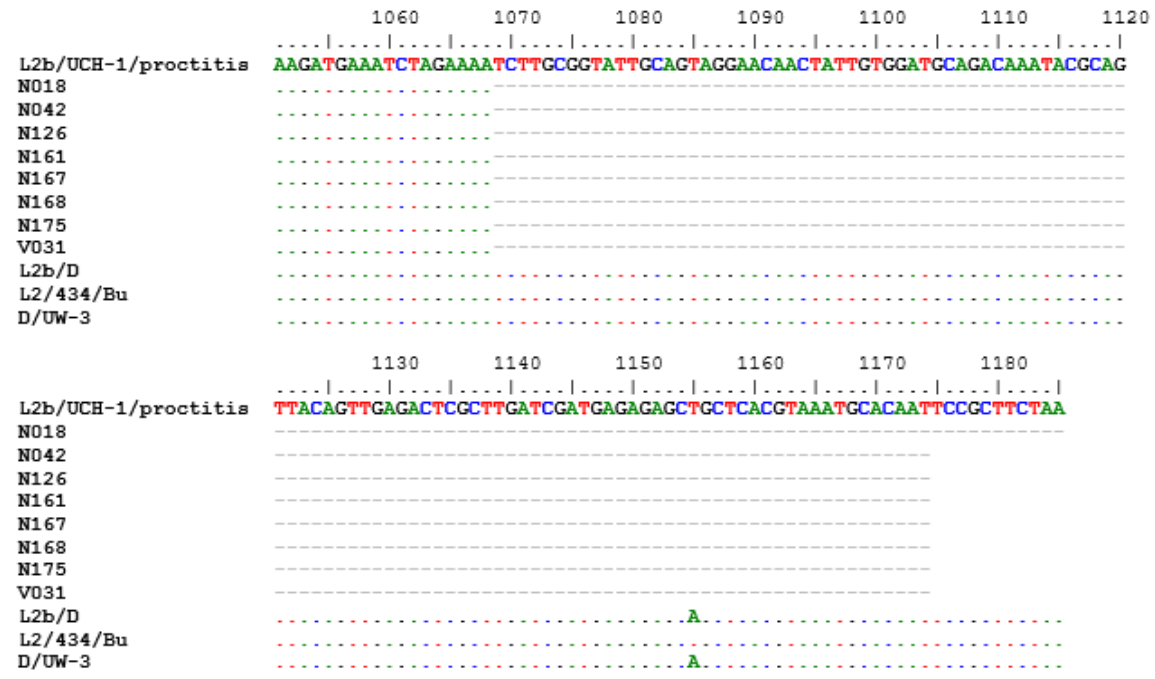

710 720 730 740 750 760 770
L2b/UCH-1/proctitis AGTTTACTATCAATAAGCCTAAAGGATATGTAGGGCAAGAATTCCCTCTTGATCTTAAAGCAGGAACAGA
N018 .....T.....A.....G.....TA.G..G..T.....C.....
N042 .....T.....A.....G.....TA.G..G..T.....C.....
N126 .....T.....A.....G.....TA.G..G..T.....C.....
N161 .....T.....A.....G.....TA.G..G..T.....C.....
N167 .....T.....A.....G.....TA.G..G..T.....C.....
N168 .....T.....A.....G.....TA.G..G..T.....C.....
N175 .....T.....A.....G.....TA.G..G..T.....C.....
V031 .....T.....A.....G.....TA.G..G..T.....C.....
L2b/D .....T.....A.....G.....TA.G..G..T.....C.....
L2/434/Bu .....
D/UW-3 .....T.....A.....G.....TA.G..G..T.....C.....


780 790 800 810 820 830 840
L2b/UCH-1/proctitis TGGTGTGACAGGAACTAAGGATGCCTCTATTGATTACCATGAATGGCAAGCAAGTTTAGCTCTCTCTTTAC
N018 ..C..C.....
N042 ..C..C.....
N126 ..C..C.....
N161 ..C..C.....
N167 ..C..C.....
N168 ..C..C.....
N175 ..C..C.....
V031 ..C..C.....
L2b/D ..C..C.....
L2/434/Bu .....
D/UW-3 ..C..C.....


850 860 870 880 890 900 910
L2b/UCH-1/proctitis AGACTGAATATGTTCACTCCCTACATTGGAGTTAAATGGTCTCGAGCAAGTTTTGATGCAGACACGATTC
N018 .....C.....C..T.....
N042 .....C.....C..T.....
N126 .....C.....C..T.....
N161 .....C.....C..T.....
N167 .....C.....C..T.....
N168 .....C.....C..T.....
N175 .....C.....C..T.....
V031 .....C.....C..T.....
L2b/D .....C.....C..T.....
L2/434/Bu .....
D/UW-3 .....C.....C..T.....


920 930 940 950 960 970 980
L2b/UCH-1/proctitis GTATTGCTCAGCCGAAGTCAGCTACAACTGTCTTTGATGTTACCACTCTGAACCCAACTATTGCTGGAGC
N018 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N042 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N126 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N161 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N167 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N168 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
N175 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
V031 ....A..C.....A..A.....G..A.T.....AC.....G..T.....
L2b/D ....A..C.....A..A.....G..A.T.....AC.....G..T.....
L2/434/Bu .....
D/UW-3 ....A..C.....A..A.....G..A.T.....AC.....G..T.....


990 1000 1010 1020 1030 1040 1050
L2b/UCH-1/proctitis TGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGATACCATGCAAATCGTTTCCTTGCAATTGAAC
N018 ..........A..G..................C..A.....................
N042 ..........A..G..................C..A.....................
N126 ..........A..G..................C..A.....................
N161 ..........A..G..................C..A.....................
N167 ..........A..G..................C..A.....................
N168 ..........A..G..................C..A.....................
N175 ..........A..G..................C..A.....................
V031 ..........A..G..................C..A.....................
L2b/D ..........A..G..................C..A.....................
L2/434/Bu ....
D/UW-3 ..........A..G..................C..A.....................


1060 1070 1080 1090 1100 1110 1120
L2b/UCH-1/proctitis AAGATGAAATCTAGAAAATCTTGCGGTATTGCAGTAGGAACAACTATTGTGGATGCAGACAAATACGCAG
N018 ....----------
N042 ....----------
N126 ....----------
N161 ....----------
N167 ....----------
N168 ....----------
N175 ....----------
V031 ....----------
L2b/D ....----------
L2/434/Bu ....----------
D/UW-3 ....----------
1130 1140 1150 1160 1170 1180
L2b/UCH-1/proctitis TTACAGTTGAGACTCGCTTGATCGATGAGAGAGCTGCTCACGTAAATGCACAATTCCGCTTCTAA
N018 ----------
N042 ----------
N126 ----------
N161 ----------
N167 ----------
N168 ----------
N175 ----------
V031 ----------
L2b/D ....A.....
L2/434/Bu ....A.....
D/UW-3 ....A.....

APPENDIX 9 Snippy Output Data

Comparing Bri038, Bri088, and Bri098 to the L2b/UCH-1/proctitis genome (AM884177)

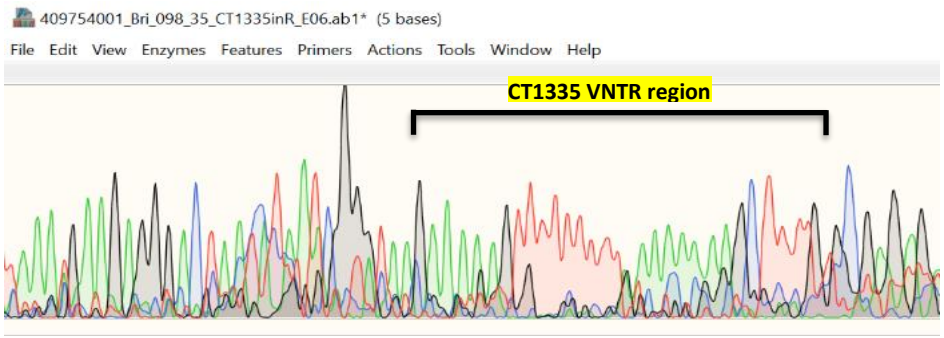
CHROM	POS	TYPE	REF	ALT	EVIDENCE	FTYPE	STRAND	NT_POS	AA_POS	LOCUS_TAG	Gene	PRODUCT
Bri038												
AM884177	59310	snp	G	T	T: 40 G: 0	CDS	-	517/1185	173/394	CTLon_0050	*ompA*	Major outer membrane protein (MOMP)
AM884177	495530	snp	G	A	A: 34 G: 0							
AM884177	630873	snp	G	A	A: 33 G: 0	CDS	+	221/735	74/244	CTLon_0529	*nqrE*	Na (+)-translocating NADH-quinone reductase subunit E
Bri088												
AM884177	59310	snp	G	T	T: 27 G: 0	CDS	-	517/1185	173/394	CTLon_0050	*ompA*	Major outer membrane protein (MOMP)
AM884177	59342	snp	C	T	T: 27 C: 0	CDS	-	485/1185	162/394	CTLon_0050	*ompA*	Major outer membrane protein (MOMP)
AM884177	481202	ins	T	TG	TG: 16 T: 0							
AM884177	495530	snp	G	A	A: 15 G: 0							
AM884177	615012	snp	T	G	G: 23 T: 0	CDS	-	851/975	284/324	CTLon_0513	*accA*	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha
AM884177	919607	ins	C	CT	CT: 17 C: 0							

CHROM	POS	TYPE	REF	ALT	EVIDENCE	FTYPE	STRAND	NT_POS	AA_POS	LOCUS_TAG	Gene	PRODUCT
Bri098												
AM884177	59342	snp	C	T	T: 53 C: 0	CDS	-	485/1185	162/394	CTLon_0050	*ompA*	Major outer membrane protein (MOMP)
AM884177	72990	snp	C	A	A: 71 C: 0	CDS	+	370/675	124/224	CTLon_0060		Conserved hypothetical protein
AM884177	480488	snp	A	G	G: 56 A: 0							
AM884177	760790	snp	G	A	A: 47 G: 0	CDS	+	1639/2076	547/691	CTLon_0640		Conserved hypothetical protein
AM884177	798316	del	AT	A	A: 60 AT: 0							
AM884177	919607	ins	C	CT	CT:49 C: 2							

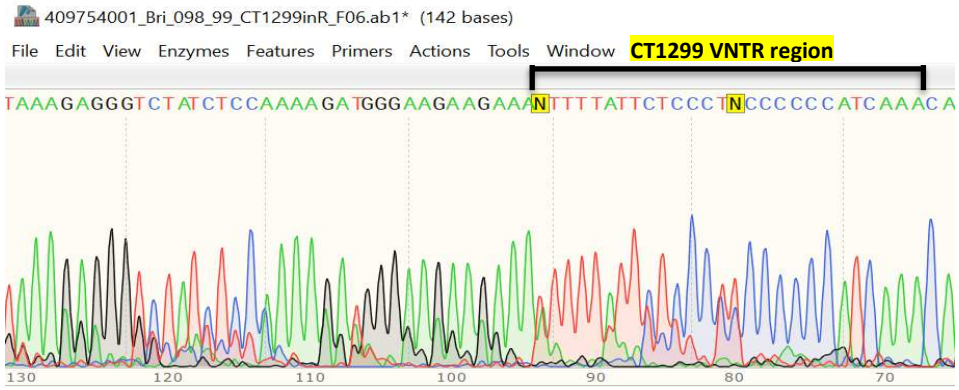
APPENDIX 10 Bri098 VNTR Sequence Chromatograms Obtained by Labiran (2014)

Sequence chromatograms for Bri098 obtained by Labiran (2014). Chromatograms show the three VNTR regions (CT1335, CT1299, and CT1291) comprising the MLVA-*ompA* genotyping system, in Bri098. The sequence chromatograms of (A) CT1335, (B) CT1299, and (C) CT1291, are shown. The VNTR regions are indicated in bold above each chromatogram. Chromatograms were viewed using SnapGene (version 5.3).

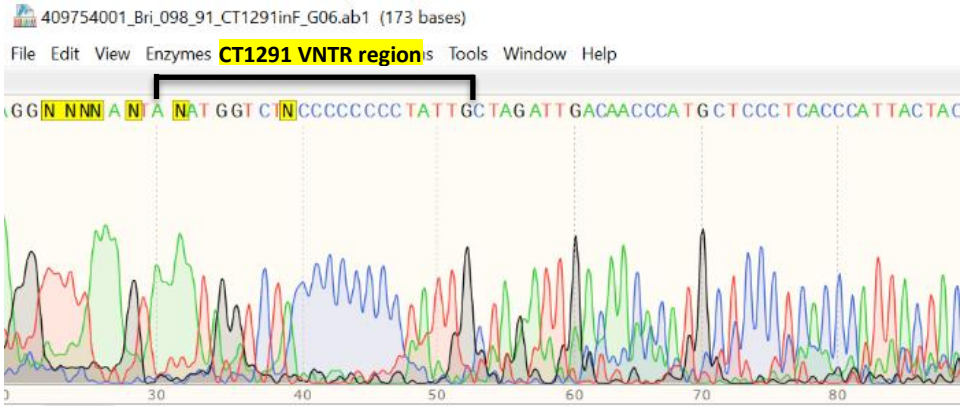
(A)



(B)

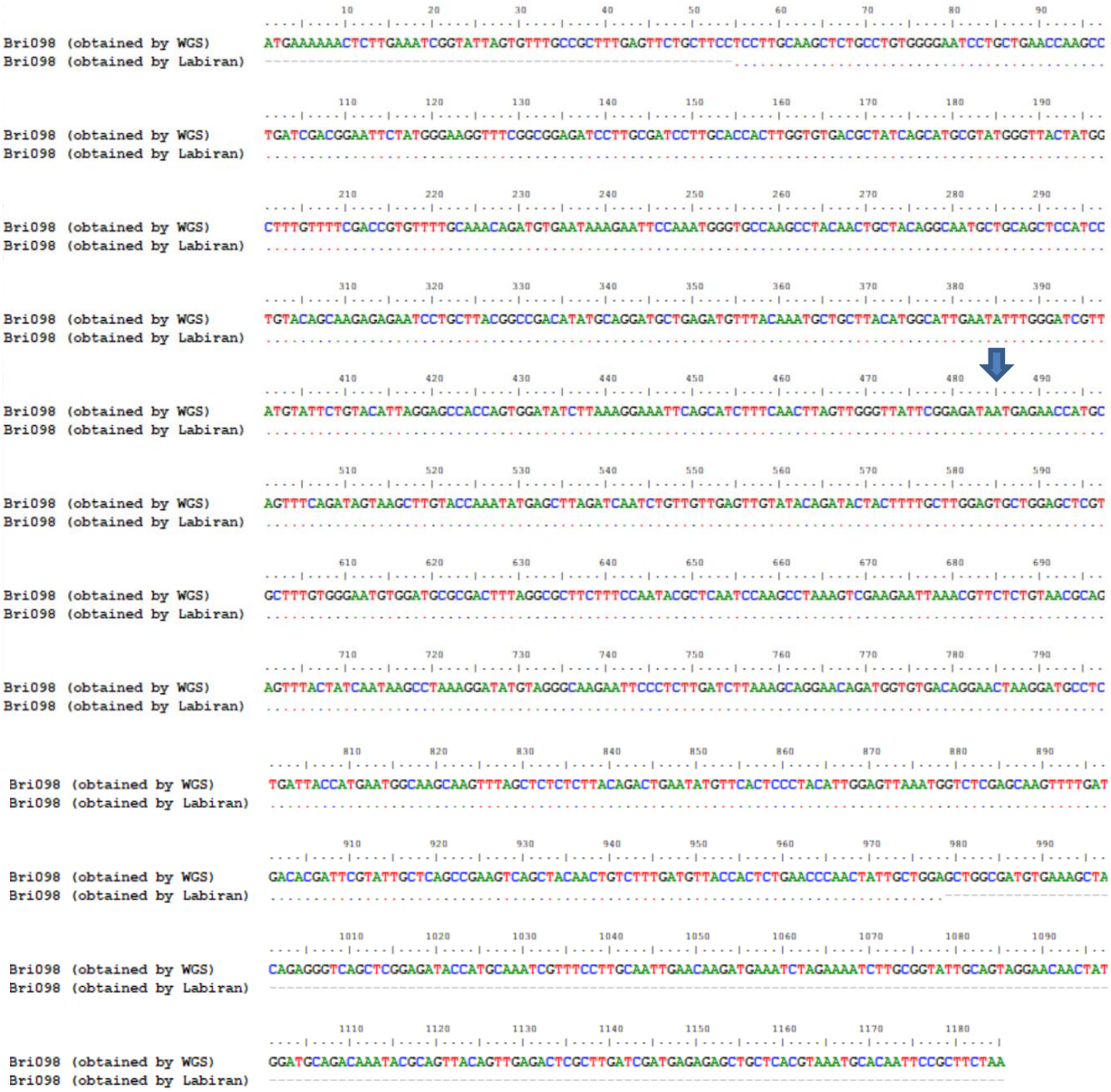


(C)



APPENDIX 11 Alignment of Bri098 *ompA* Sequences

Sequence alignment of *ompA* from Bri098. Included in the alignment is the 923 bp *ompA* sequence obtained for Bri098 by Labiran (2014), and the 1185 bp *ompA* sequence obtained for Bri098 in Chapter 6 by whole genome sequencing (WGS). Coloured dots indicate bases that were identical to *ompA* obtained by WGS. Sequences were aligned using ClustalW in BioEdit (version 7.0.5.3). The G485A SNP that reverts the Bri098 *ompA* sequence from L2b to L2 is present in both *ompA* sequences and is indicated by an arrow at nucleotide position 485.



LIST OF REFERENCES

van Aar, F. *et al.* (2020) 'Increasing trends of lymphogranuloma venereum among HIV-negative and asymptomatic men who have sex with men, the Netherlands, 2011 to 2017', *Eurosurveillance*, 25(14). doi: 10.2807/1560-7917.ES.2020.25.14.1900377.

Abdelrahman, Y. M. and Belland, R. J. (2005) 'The chlamydial developmental cycle.', *FEMS microbiology reviews*, 29(5), pp. 949–59. doi: 10.1016/j.femsre.2005.03.002.

Afgan, E., et al. (2018) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.', *Nucleic Acids Res.,* 46(W1):W537-W544. doi: 10.1093/nar/gky379.

Alexander, S., Martin, I. and Ison, C. (2007) 'Confirming the Chlamydia trachomatis status of referred rectal specimens.', *Sexually transmitted infections*, 83(4), pp. 327–9. doi: 10.1136/sti.2006.024620.

Di Altobrando, A. *et al.* (2019) 'Lymphogranuloma venereum proctitis mimicking inflammatory bowel diseases in 11 patients: a 4-year single-center experience', *Crohn's & Colitis 360*, 1(1). doi: 10.1093/crocol/otz004.

Andersson, P. *et al.* (2016) 'Chlamydia trachomatis from Australian Aboriginal people with trachoma are polyphyletic composed of multiple distinctive lineages', *Nature Communications*, 7(1), p. 10688. doi: 10.1038/ncomms10688.

Aral, S. O. (2000) 'Patterns of sexual mixing: mechanisms for or limits to the spread of STIs?', *Sexually Transmitted Infections*, 76(6), pp. 415–416. doi: 10.1136/sti.76.6.415.

Bankevich, A. *et al.* (2012) 'SPAdes: a new genome assembly algorith and its applications to single-cell sequencing', *J Comput Biol*, 5(19), pp. 455–77.

Bannerjee, A. and Nelson, D. E. (2019) 'How Chlamydia trachomatis conquered gut microbiome-derived antimicrobial compounds and found a new home in the eye', *Proc Natl Acad Sci USA,* 116(25):12136-12138. doi: 10.1073/pnas.1907647116.

Barron, A. L. *et al.* (1981) 'A new animal model for the study of Chlamydia trachomatis genital infections: infection of mice with the agent of mouse pneumonitis.', *The Journal of infectious diseases*, 143(1), pp. 63–6. doi: 10.1093/infdis/143.1.63.

Bauwens, J. E. *et al.* (1995) 'Infection with Chlamydia trachomatis lymphogranuloma venereum serovar L1 in homosexual men with proctitis: molecular analysis of an unusual case cluster.',

Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 20(3), pp. 576–81. doi: 10.1093/clinids/20.3.576.

Bavoil, P., Ohlin, A. and Schachter, J. (1984) 'Role of disulfide bonding in outer membrane structure and permeability in Chlamydia trachomatis.', *Infection and immunity*, 44(2), pp. 479–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6715046.

Beatty, W. L. *et al.* (1994) 'Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence.', *Infection and immunity*, 62(9), pp. 3705–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8063385.

Beatty, W. L., Morrison, R. P. and Byrne, G. I. (1995) 'Reactivation of persistent Chlamydia trachomatis infection in cell culture.', *Infection and immunity*, 63(1), pp. 199–205. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7806358.

van Belkum, A. *et al.* (2007) 'Guidelines for the validation and application of typing methods for use in bacterial epidemiology.', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 13 Suppl 3, pp. 1–46. doi: 10.1111/j.1469-0691.2007.01786.x.

Belland, R. J. *et al.* (2001) 'Chlamydia trachomatis cytotoxicity associated with complete and partial cytotoxin genes', *Proceedings of the National Academy of Sciences*, 98(24), pp. 13984–13989. doi: 10.1073/pnas.241377698.

Berhane, Y., Worku, A. and Bejiga, A. (2006) 'National Survey on Blindness, Low Vision and Trachoma in Ethiopia', 21.

Bolger, A., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120.

Bom, R. J. M. *et al.* (2011) 'Evaluation of high-resolution typing methods for Chlamydia trachomatis in samples from heterosexual couples.', *Journal of clinical microbiology*, 49(8), pp. 2844–53. doi: 10.1128/JCM.00128-11.

Bommana, S. *et al.* (2021) 'Tryptophan Operon Diversity Reveals Evolutionary Trends among Geographically Disparate Chlamydia trachomatis Ocular and Urogenital Strains Affecting Tryptophan Repressor and Synthase Function', *mBio*. Edited by N. R. Salama, 12(3). doi: 10.1128/mBio.00605-21.

Borges, V. *et al.* (2019) 'Chlamydia trachomatis: When the virulence-associated genome backbone imports a prevalence-associated major antigen signature', *Microbial Genomics*. Microbiology

Society, 5(11). doi: 10.1099/mgen.0.000313.

Borges, V. and Gomes, J. P. (2015a) 'Deep comparative genomics among Chlamydia trachomatis lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 32, pp. 74–88. doi: 10.1016/j.meegid.2015.02.026.

Borges, V. and Gomes, J. P. (2015b) 'Deep comparative genomics among Chlamydia trachomatis lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation', *Infection, Genetics and Evolution*, 32, pp. 74–88. doi: 10.1016/j.meegid.2015.02.026.

Brownell, A. D. *et al.* (2020) 'Chlamydia Conjunctivitis in 2 Prepubertal Children: An Uncommon Presentation of Child Sexual Abuse.', *Pediatric emergency care*, 36(8), pp. e473–e475. doi: 10.1097/PEC.0000000000001449.

Brunelle, B. W. and Sensabaugh, G. F. (2006) 'The ompA gene in Chlamydia trachomatis differs in phylogeny and rate of evolution from other regions of the genome.', *Infection and immunity*, 74(1), pp. 578–85. doi: 10.1128/IAI.74.1.578-585.2006.

Brunelle, B. W. and Sensabaugh, G. F. (2012) 'Nucleotide and phylogenetic analyses of the Chlamydia trachomatis ompA gene indicates it is a hotspot for mutation.', *BMC research notes*, 5, p. 53. doi: 10.1186/1756-0500-5-53.

Burton, M. J. (2007) 'Trachoma: an overview', *British Medical Bulletin*, 84(1), pp. 99–116. doi: 10.1093/bmb/ldm034.

Burton, M. J. and Mabey, D. C. W. (2009) 'The Global Burden of Trachoma: A Review', *PLoS Neglected Tropical Diseases*. Edited by S. Brooker, 3(10), p. e460. doi: 10.1371/journal.pntd.0000460.

Byrne, G. I. (2010) 'Chlamydia trachomatis Strains and Virulence: Rethinking Links to Infection Prevalence and Disease Severity', *The Journal of Infectious Diseases*, 201(S2), pp. 126–133. doi: 10.1086/652398.

Caldwell, H. D. *et al.* (2003) 'Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates.', *The Journal of clinical investigation*, 111(11), pp. 1757–69. doi: 10.1172/JCI17993.

Caldwell, H. D., Kromhout, J. and Schachter, J. (1981) 'Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis.', *Infection and immunity*, 31(3), pp. 1161–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7228399.

Carver, T. *et al.* (2012) 'Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data', *Bioinformatics*, 28(4), pp. 464–469.

Ceovic, R. and Gulin, S. J. (2015) 'Lymphogranuloma venereum: diagnostic and treatment challenges.', *Infection and drug resistance*, 8, pp. 39–47. doi: 10.2147/IDR.S57540.

Chen, C.-Y. *et al.* (2007) 'The Molecular Diagnosis of Lymphogranuloma Venereum: Evaluation of a Real-Time Multiplex Polymerase Chain Reaction Test Using Rectal and Urethral Specimens', *Sexually Transmitted Diseases*, 34(7), pp. 451–455. doi: 10.1097/01.olq.0000245957.02939.ea.

Christerson, L. *et al.* (2010) 'Typing of Lymphogranuloma Venereum Chlamydia trachomatis Strains', *Emerging Infectious Diseases*, 16(11), pp. 1777–1779. doi: 10.3201/eid1611.100379.

Cole, M. J. *et al.* (2020a) 'Substantial underdiagnosis of lymphogranuloma venereum in men who have sex with men in Europe: preliminary findings from a multicentre surveillance pilot.', *Sexually transmitted infections*, 96(2), pp. 137–142. doi: 10.1136/sextrans-2019-053972.

Cole, M. J. *et al.* (2020b) 'Substantial underdiagnosis of lymphogranuloma venereum in men who have sex with men in Europe: preliminary findings from a multicentre surveillance pilot', *Sexually Transmitted Infections*, 96(2), pp. 137–142. doi: 10.1136/sextrans-2019-053972.

Cossé, M. M., Hayward, R. D. and Subtil, A. (2016) 'One Face of Chlamydia trachomatis: The Infectious Elementary Body', in, pp. 35–58. doi: 10.1007/82_2016_12.

Dai, W. and Li, Z. (2014) 'Conserved type III secretion system exerts important roles in C. trachomatis', *Int J Clin Exp Pathol*, 7(9), pp. 5404–14.

Darville, T. (2005) 'Chlamydia trachomatis Infections in Neonates and Young Children', *Seminars in Pediatric Infectious Diseases*, 16(4), pp. 235–244. doi: 10.1053/j.spid.2005.06.004.

Dean, D. *et al.* (1992) 'Comparison of the Major Outer Membrane Protein Variant Sequence Regions-of B/Ba Isolates: A Molecular Epidemiologic Approach to Chlamydia trachomatis Infections', *Journal of Infectious Diseases*, 166(2), pp. 383–392. doi: 10.1093/infdis/166.2.383.

Dean, D. *et al.* (2009) 'Predicting phenotype and emerging strains among Chlamydia trachomatis infections.', *Emerging infectious diseases*, 15(9), pp. 1385–94. doi: 10.3201/eid1509.090272.

Dosekun, O. *et al.* (2013) 'Lymphogranuloma venereum detected from the pharynx in four London men who have sex with men.', *International journal of STD & AIDS*, 24(6), pp. 495–6. doi: 10.1177/0956462412472830.

Elwell, C., Mirrashidi, K. and Engel, J. (2016) 'Chlamydia cell biology and pathogenesis', *Nature*

Reviews Microbiology, 14(6), pp. 385–400. doi: 10.1038/nrmicro.2016.30.

Emerson, P. M. *et al.* (2000) 'Transmission ecology of the fly Musca sorbens, a putative vector of trachoma', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(1), pp. 28–32. doi: 10.1016/S0035-9203(00)90427-9.

Emerson, P. M. *et al.* (2008) 'Integrating an NTD with One of "The Big Three": Combined Malaria and Trachoma Survey in Amhara Region of Ethiopia', *PLoS Neglected Tropical Diseases*. Edited by J. Schachter, 2(3), p. e197. doi: 10.1371/journal.pntd.0000197.

European Centre for Disease Control (2019) *Lymphogranuloma Venereum: Annual Epidemiological Report for 2018*. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/lymphogranuloma-venereum-aer-2018.pdf.

Everett, K. D., Bush, R. M. and Andersen, A. A. (1999) 'Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards', *International journal of systematic bacteriology*, 49 Pt 2, pp. 415–40. doi: 10.1099/00207713-49-2-415.

Feher, V. A. *et al.* (2013) 'A 3-Dimensional Trimeric β-Barrel Model for Chlamydia MOMP Contains Conserved and Novel Elements of Gram-Negative Bacterial Porins', *PLoS ONE*. Edited by A. Motta, 8(7), p. e68934. doi: 10.1371/journal.pone.0068934.

Fehlner-Gardiner, C. *et al.* (2002) 'Molecular basis defining human Chlamydia trachomatis tissue tropism. A possible role for tryptophan synthase.', *The Journal of biological chemistry*, 277(30), pp. 26893–903. doi: 10.1074/jbc.M203937200.

Fukushi, H. and Hirai, K. (1992) 'Proposal of Chlamydia pecorum sp. nov. for Chlamydia strains derived from ruminants.', *International journal of systematic bacteriology*, 42(2), pp. 306–8. doi: 10.1099/00207713-42-2-306.

Furegato, M. *et al.* (2016) 'Examining the role of socioeconomic deprivation in ethnic differences in sexually transmitted infection diagnosis rates in England: evidence from surveillance data.', *Epidemiology and infection*, 144(15), pp. 3253–3262. doi: 10.1017/S0950268816001679.

Gafos, M. *et al.* (2019) 'The Context of Sexual Risk Behaviour Among Men Who Have Sex with Men Seeking PrEP, and the Impact of PrEP on Sexual Behaviour', *AIDS and Behavior*, 23(7), pp. 1708–1720. doi: 10.1007/s10461-018-2300-5.

Garza, A., et al. (2021) 'Homeostasis and defense at the surface of the eye. The conjunctival microbiota.', *Current Eye Research,* 46(1): pp: 1-6. doi: 10.1080/02713683.20201788100.

Gehre, L. *et al.* (2016) 'Sequestration of host metabolism by an intracellular pathogen.', *eLife*, 5, p. e12552. doi: 10.7554/eLife.12552.

Geisler, W. M. *et al.* (2002) 'Epidemiology of anorectal chlamydial and gonococcal infections among men having sex with men in Seattle: utilizing serovar and auxotype strain typing.', *Sexually transmitted diseases*, 29(4), pp. 189–95. doi: 10.1097/00007435-200204000-00001.

Gibbs, R. S. (1987) 'Microbiology of the female genital tract.', *American journal of obstetrics and gynecology*, 156(2), pp. 491–5. doi: 10.1016/0002-9378(87)90318-8.

Giffard, P. M. *et al.* (2016) 'Chlamydia trachomatis genotypes in a cross-sectional study of urogenital samples from remote Northern and Central Australia', *BMJ Open*, 6(1), p. e009624. doi: 10.1136/bmjopen-2015-009624.

Giffard, P. M., et al. (2018) 'Contaminated fingers: a potential cause of Chlamydia trachomatis-positive urine specimens.', *Sex Trans Infect,* 94(1):32-36. doi: 10.1136/sextrans-2016-053081.

Gilbart, V. L. *et al.* (2013) 'High-risk drug practices in men who have sex with men', *The Lancet*, 381(9875), pp. 1358–1359. doi: 10.1016/S0140-6736(13)60882-X.

Gilbart, V. L. *et al.* (2015) 'Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with Shigella flexneri 3a in England and Wales.', *Sexually transmitted infections*, 91(8), pp. 598–602. doi: 10.1136/sextrans-2015-052014.

Gomes, J. P. *et al.* (2006) 'Polymorphisms in the nine polymorphic membrane proteins of Chlamydia trachomatis across all serovars: evidence for serovar Da recombination and correlation with tissue tropism.', *Journal of bacteriology*, 188(1), pp. 275–86. doi: 10.1128/JB.188.1.275-286.2006.

Gomes, J. P. *et al.* (2007) 'Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots.', *Genome research*, 17(1), pp. 50–60. doi: 10.1101/gr.5674706.

Götz, H. M. *et al.* (2004) '[A cluster of lymphogranuloma venereum among homosexual men in Rotterdam with implications for other countries in Western Europe].', *Nederlands tijdschrift voor geneeskunde*, 148(9), pp. 441–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15038207.

Gravningen, K. *et al.* (2012) 'Multilocus sequence typing of genital Chlamydia trachomatis in

Norway reveals multiple new sequence types and a large genetic diversity.', *PloS one*, 7(3), p. e34452. doi: 10.1371/journal.pone.0034452.

Grayston, J. T. et al. (1960) 'Trachoma-studies of etiology, laboratory diagnosis, and prevention', *J Am Med Assoc,* 172: pp. 1577-86. doi: 10.1001/jama.1960.03020150001001.

Grayston, J. T. *et al.* (1989) 'Chlamydia pneumoniae sp. nov. for Chlamydia sp. Strain TWAR', *International Journal of Systematic Bacteriology*, 39(1), pp. 88–90. doi: 10.1099/00207713-39-1-88.

Greub, G. (2010) 'International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of the Chlamydiae: minutes of the inaugural closed meeting, 21 March 2009, Little Rock, AR, USA.', *International journal of systematic and evolutionary microbiology*, 60(Pt 11), pp. 2691–2693. doi: 10.1099/ijs.0.028225-0.

Grov, C. *et al.* (2015) 'HIV Serosorting, Status Disclosure, and Strategic Positioning Among Highly Sexually Active Gay and Bisexual Men', *AIDS Patient Care and STDs*, 29(10), pp. 559–568. doi: 10.1089/apc.2015.0126.

Guo, W. *et al.* (2016) 'Chlamydia gallinacea, not C. psittaci, is the endemic chlamydial species in chicken (Gallus gallus)', *Scientific Reports*, 6(1), p. 19638. doi: 10.1038/srep19638.

Hackstadt, T., Todd, W. J. and Caldwell, H. D. (1985) 'Disulfide-mediated interactions of the chlamydial major outer membrane protein: role in the differentiation of chlamydiae?', *Journal of Bacteriology*, 161(1), pp. 25–31. doi: 10.1128/JB.161.1.25-31.1985.

Hadfield, J. *et al.* (2017) 'Comprehensive global genome dynamics of Chlamydia trachomatis show ancient diversification followed by contemporary mixing and recent lineage expansion', *Genome Research*, 27(7), pp. 1220–1229. doi: 10.1101/gr.212647.116.

Halberstadter, L. and Von Prowazek, S. (1907) 'Uber Zelleinschlusse parasitarer Natur beim Trachom', *Arb K Gesundh-Amte*, pp. 44–47.

Halberstadter, L. and Von Prowazek, S. (1910) 'Ueber die Bedeutung der Chlamydozoen bei Trachom und Blennorhoea', *Berl Klin Wochenschr*, 47, p. 661.

Halioua, B. *et al.* (no date) 'Ano-rectal lymphogranuloma venereum: 22 cases reported in a sexually transmited infections center in Paris.', *European journal of dermatology : EJD*, 16(2), pp. 177–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16581572.

Hamlyn, E. and Taylor, C. (2006) 'Sexually transmitted proctitis.', *Postgraduate medical journal*,

82(973), pp. 733–6. doi: 10.1136/pmj.2006.048488.

Hammerschlag, M. R. (1998) 'Sexually transmitted diseases in sexually abused children: medical and legal implications.', *Sexually transmitted infections*, 74(3), pp. 167–74. doi: 10.1136/sti.74.3.167.

Hammerschlag, M. R. (2011) 'Chlamydial and gonococcal infections in infants and children.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 53 Suppl 3, pp. S99-102. doi: 10.1093/cid/cir699.

Hanahan, D. (1983) 'Studies on transformation of Escherichia coli with plasmids', *Journal of Molecular Biology*, 166(4), pp. 557–580. doi: 10.1016/S0022-2836(83)80284-8.

Harris, S. R. *et al.* (2012) 'Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing.', *Nature genetics*, 44(4), pp. 413–9, S1. doi: 10.1038/ng.2214.

Herrmann, B. *et al.* (2015) 'Global Multilocus Sequence Type Analysis of Chlamydia trachomatis Strains from 16 Countries.', *Journal of clinical microbiology*, 53(7), pp. 2172–9. doi: 10.1128/JCM.00249-15.

Hillis, D. M. and Bull, J. J. (1993) 'An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis', *Systematic Biology*, 42(2), pp. 182–192. doi: 10.1093/sysbio/42.2.182.

Hoenderboom, B. M. *et al.* (2019) 'Relation between Chlamydia trachomatis infection and pelvic inflammatory disease, ectopic pregnancy and tubal factor infertility in a Dutch cohort of women previously tested for chlamydia in a chlamydia screening trial.', *Sexually transmitted infections*, 95(4), pp. 300–306. doi: 10.1136/sextrans-2018-053778.

Hogan, R. J. *et al.* (2004) 'Chlamydial persistence: beyond the biphasic paradigm.', *Infection and immunity*, 72(4), pp. 1843–55. doi: 10.1128/iai.72.4.1843-1855.2004.

Hologic (2017) *Aptima Combo 2 Assay (Panther System)*. Available at: https://www.hologic.com/sites/default/files/package-insert/502446-IFU-PI_003_01.pdf.

Hotinger, J. A., Pendergrass, H. A. and May, A. E. (2021) 'Molecular Targets and Strategies for Inhibition of the Bacterial Type III Secretion System (T3SS); Inhibitors Directly Binding to T3SS Components', *Biomolecules*, 11(2), p. 316. doi: 10.3390/biom11020316.

Hughes, G. *et al.* (2013) 'Lymphogranuloma venereum diagnoses among men who have sex with

men in the U.K.: interpreting a cross-sectional study using an epidemic phase-specific framework.', *Sexually transmitted infections*, 89(7), pp. 542–7. doi: 10.1136/sextrans-2013-051051.

Hunter, P. R. and Gaston, M. A. (1988) 'Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity.', *Journal of clinical microbiology*, 26(11), pp. 2465–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3069867.

Hybiske, K. and Stephens, R. S. (2007) 'Mechanisms of host cell exit by the intracellular bacterium Chlamydia', *Proceedings of the National Academy of Sciences*, 104(27), pp. 11430–11435. doi: 10.1073/pnas.0703218104.

Ikryannikova, L. N. *et al.* (2010) 'Comparative evaluation of new typing schemes for urogenital Chlamydia trachomatis isolates.', *FEMS immunology and medical microbiology*, 59(2), pp. 188–96. doi: 10.1111/j.1574-695X.2010.00678.x.

Ingram, D. L. *et al.* (1992) 'Epidemiology of adult sexually transmitted disease agents in children being evaluated for sexual abuse', *The Pediatric Infectious Disease Journal*, 11(11), pp. 945–950. doi: 10.1097/00006454-199211110-00008.

Isaksson, J. *et al.* (2017) 'Lymphogranuloma venereum rates increased and Chlamydia trachomatis genotypes changed among men who have sex with men in Sweden 2004-2016.', *Journal of medical microbiology*, 66(11), pp. 1684–1687. doi: 10.1099/jmm.0.000597.

Jacobson, D. L. *et al.* (2000) 'Histologic development of cervical ectopy: relationship to reproductive hormones.', *Sexually transmitted diseases*, 27(5), pp. 252–8. doi: 10.1097/00007435-200005000-00003.

Janssen, K. H., et al. (2019) 'Assessment of rectal Chlamydia trachomatis visible load in women by viability-PCR.', *Sex Trans Infect,* 96(2):85-88. doi: 10.1136/sextrans-2019-054002.

Jeffrey, B. M. *et al.* (2010) 'Genome sequencing of recent clinical Chlamydia trachomatis strains identifies loci associated with tissue tropism and regions of apparent recombination.', *Infection and immunity*, 78(6), pp. 2544–53. doi: 10.1128/IAI.01324-09.

Joseph Davey, D., Medline, A. and Klausner, J. D. (2016) 'Screening pregnant women in the 2015 European guideline on the management of Chlamydia trachomatis infections.', *International journal of STD & AIDS*, 27(12), pp. 1134–1136. doi: 10.1177/0956462416650600.

Joseph, S. *et al.* (2012) 'Population Genomics of Chlamydia trachomatis: Insights on Drift, Selection, Recombination, and Population Structure', *Mol Biol Evol*, 29(12), pp. 3933–3946.

Joseph, S. J. *et al.* (2011) ‘Interplay of recombination and selection in the genomes of Chlamydia trachomatis.’, *Biology direct*, 6, p. 28. doi: 10.1186/1745-6150-6-28.

Kari, L. et al. (2008) 'Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations.', *J Infect Dis,* 197(3): pp449-56. doi: 10.1086/525285.

Kestelyn, P. (2008) ‘Adult Inclusion Conjunctivitis’, in *Cornea and External Eye Disease*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 179–184. doi: 10.1007/978-3-540-33681-5_9.

Kirk, E. *et al.* (2008) ‘Chlamydia trachomatis infection in patients attending an Early Pregnancy Unit: prevalence, symptoms, pregnancy location and viability.’, *Acta obstetricia et gynecologica Scandinavica*, 87(6), pp. 601–7. doi: 10.1080/00016340802056095.

Klint, M. *et al.* (2007) ‘High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis.’, *Journal of clinical microbiology*, 45(5), pp. 1410–4. doi: 10.1128/JCM.02301-06.

Korhonen, S., Hiltunen-Back, E. and Puolakkainen, M. (2012) ‘Genotyping of Chlamydia trachomatis in rectal and pharyngeal specimens: identification of LGV genotypes in Finland.’, *Sexually transmitted infections*, 88(6), pp. 465–9. doi: 10.1136/sextrans-2011-050458.

Kropp, R. Y., Wong, T. and Canadian LGV Working Group (2005) ‘Emergence of lymphogranuloma venereum in Canada.’, *CMAJ : Canadian Medical Association journal = journal de l’Association medicale canadienne*, 172(13), pp. 1674–6. doi: 10.1503/cmaj.050621.

Labiran, C. *et al.* (2012) ‘Genotyping markers used for multi locus VNTR analysis with ompA (MLVA-ompA) and multi sequence typing (MST) retain stability in Chlamydia trachomatis.’, *Frontiers in cellular and infection microbiology*, 2, p. 68. doi: 10.3389/fcimb.2012.00068.

Labiran, C. (2014) *Molecular epidemiology of Chlamydia trachomatis: Evaluation, implementation and development of high resolution genotyping*. Available at: https://eprints.soton.ac.uk/374912/1/Clare%2520Labiran_%2520Thesis_%252010_June_14.pdf.

Labiran, C. *et al.* (2016) ‘Highly diverse MLVA-ompA genotypes of rectal Chlamydia trachomatis among men who have sex with men in Brighton, UK and evidence for an HIV-related sexual network.’, *Sexually transmitted infections*, 92(4), pp. 299–304. doi: 10.1136/sextrans-2015-052261.

Labiran, C. *et al.* (2017) ‘Detailed molecular epidemiology of Chlamydia trachomatis in the population of Southampton attending the genitourinary medicine clinic in 2012-13 reveals the

presence of long established genotypes and transitory sexual networks.', *PloS one*, 12(9), p. e0185059. doi: 10.1371/journal.pone.0185059.

Lambden, P. R. et al. (2006) 'The effect of penicillin on Chlamydia trachomatis DNA replication', *Microbiology*, 152(Pt 9):2573-2578. doi: 10.1099/mic.0.29032-0.

Lan, J. *et al.* (1994) 'Improved PCR sensitivity for direct genotyping of Chlamydia trachomatis serovars by using a nested PCR.', *Journal of clinical microbiology*, 32(2), pp. 528–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8150967.

Lanjouw, E. *et al.* (2016) '2015 European guideline on the management of Chlamydia trachomatis infections.', *International journal of STD & AIDS*, 27(5), pp. 333–48. doi: 10.1177/0956462415618837.

Larsen, B. and Monif, G. R. (2001) 'Understanding the bacterial flora of the female genital tract.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 32(4), pp. e69-77. doi: 10.1086/318710.

Last, A. R., et al. (2018) 'Population-based analysis of ocular Chlamydia trachomatis in trachoma-endemic West African communities identifies genomic markers of disease severity.', *Genome Med,* 10(1):15. doi: 10.1186/s13073-018-0521-x.

Lee, J. K. *et al.* (2018) 'Replication-dependent size reduction precedes differentiation in Chlamydia trachomatis.', *Nature communications*, 9(1), p. 45. doi: 10.1038/s41467-017-02432-0.

Levine, J. S., Smith, P. D. and Brugge, W. R. (1980) 'Chronic proctitis in male homosexuals due to lymphogranuloma venereum.', *Gastroenterology*, 79(3), pp. 563–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7429118.

Lewis, N. et al. (2012) 'Chlamydia and gonorrhoea contamination of clinic surfaces.', *Sex Trans Infect,* 88(6):418-21. doi: 10.1136/sextrans-2012-050543.

Liang, P. *et al.* (2018) 'Dynamic energy dependency of Chlamydia trachomatis on host cell metabolism during intracellular growth: Role of sodium-based energetics in chlamydial ATP generation', *Journal of Biological Chemistry*, 293(2), pp. 510–522. doi: 10.1074/jbc.M117.797209.

Lindner, K. (1909) 'Zur Trachomforschung', *Z. Augenheilkd*, 22, pp. 547–549.

Lindner, K. (1910) 'Zur Atiologie der gonokokkenfreien Urethritis', *Wiener klinische Wochenschrift*, 23, p. 283.

Lutter, E. I. *et al.* (2010) 'Phylogenetic analysis of Chlamydia trachomatis Tarp and correlation with

clinical phenotype.', *Infection and immunity*, 78(9), pp. 3678–88. doi: 10.1128/IAI.00515-10.

Mabey, D. (2002) 'Lymphogranuloma venereum', *Sexually Transmitted Infections*, 78(2), pp. 90–92. doi: 10.1136/sti.78.2.90.

Macdonald, N. *et al.* (2005) 'Initial results of enhanced surveillance for lymphogranuloma venereum (LGV) in England.', *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 10(1), p. E050127.5. doi: 10.2807/esw.10.04.02629-en.

Maiden, M. C. J. (2006) 'Multilocus sequence typing of bacteria.', *Annual review of microbiology*, 60, pp. 561–88. doi: 10.1146/annurev.micro.59.030804.121325.

Manning, C. *et al.* (2021) 'High-resolution genotyping of Lymphogranuloma Venereum (LGV) strains of Chlamydia trachomatis in London using multi-locus VNTR analysis-ompA genotyping (MLVA-ompA).', *PloS one*, 16(7), p. e0254233. doi: 10.1371/journal.pone.0254233.

Marangoni, A. *et al.* (2021) 'Lymphogranuloma venereum genovariants in men having sex with men in Italy', *Sexually Transmitted Infections*, 97(6), pp. 441–445. doi: 10.1136/sextrans-2020-054700.

Martin-Iguacel, R. *et al.* (2010) 'Lymphogranuloma venereum proctocolitis: a silent endemic disease in men who have sex with men in industrialised countries', *European Journal of Clinical Microbiology & Infectious Diseases*, 29(8), pp. 917–925. doi: 10.1007/s10096-010-0959-2.

McComb, D. E., and Bell Jr., S. D. (1967) 'The application of an in vitro typing test to TRIC Bedsoniae.', *Am J Ophthalmol,* 63(5), pp. 1429-37. doi: 10.1016/0002-9394(67)94127-x.

McCormack, S. *et al.* (2016) 'Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial', *The Lancet*, 387(10013), pp. 53–60. doi: 10.1016/S0140-6736(15)00056-2.

McDaid, L. M. and Hart, G. J. (2010) 'Sexual risk behaviour for transmission of HIV in men who have sex with men: recent findings and potential interventions.', *Current opinion in HIV and AIDS*, 5(4), pp. 311–5. doi: 10.1097/COH.0b013e32833a0b86.

McDaid, L. M. and Hart, G. J. (2012) 'Serosorting and Strategic Positioning During Unprotected Anal Intercourse', *Sexually Transmitted Diseases*, 39(9), pp. 735–738. doi: 10.1097/OLQ.0b013e31825a3a3c.

Mercer, C. H. *et al.* (2013) 'Changes in sexual attitudes and lifestyles in Britain through the life

course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal).', *Lancet (London, England)*, 382(9907), pp. 1781–94. doi: 10.1016/S0140-6736(13)62035-8.

Merck (2020) *Blue-White Screening & Protocols for Colony Selection*. Available at: https://www.sigmaaldrich.com/technical-documents/articles/biology/blue-white-screening.html (Accessed: 12 November 2020).

Messing, J. (1979) 'A multipurpose cloning system based on the single-stranded DNA bacteriophage M13', *Recombinant DNA Technical Bulletin*, 2(2), pp. 43–48.

Meyer, T. (2016) 'Diagnostic Procedures to Detect Chlamydia trachomatis Infections', *Microorganisms*, 4(3), p. 25. doi: 10.3390/microorganisms4030025.

Miles, E. W. (2001) 'Tryptophan synthase: a multienzyme complex with an intramolecular tunnel.', *Chemical record (New York, N.Y.)*, 1(2), pp. 140–51. doi: 10.1002/tcr.4.

Miles, E. W. (2013) 'The tryptophan synthase α2β2 complex: a model for substrate channeling, allosteric communication, and pyridoxal phosphate catalysis.', *The Journal of biological chemistry*, 288(14), pp. 10084–91. doi: 10.1074/jbc.X113.463331.

Mitchell, A. *et al.* (2021) 'Reducing suspicion of sexual abuse in paediatric chlamydial conjunctivitis using ompA genotyping.', *BMJ case reports*, 14(3). doi: 10.1136/bcr-2020-238871.

Miyagawa, Y. *et al.* (1935) 'Fourth report: studies on the virus lymphogranuloma inguinale Nicolas, Favre and Durand. Cultivation of the virus on the chrorioallantoic membrane of the chicken embryo', *Jpn J Exp Med*, 13, pp. 733–738.

Miyairi, I. *et al.* (2006) 'Different Growth Rates of Chlamydia trachomatis Biovars Reflect Pathotype', *The Journal of Infectious Diseases*, 194(3), pp. 350–357. doi: 10.1086/505432.

Miyairi, I. et al. (2011) 'Chlamydia psittaci genetic variants differ in virulence by modulation of host immunity.', *J Infect Dis,* 204(4):654-63. doi: 10.1093/infdis/jir333.

Moelleken, K. and Hegemann, J. H. (2008) 'The Chlamydia outer membrane protein OmcB is required for adhesion and exhibits biovar-specific differences in glycosaminoglycan binding.', *Molecular microbiology*, 67(2), pp. 403–19. doi: 10.1111/j.1365-2958.2007.06050.x.

Morré, S. A. *et al.* (1998) 'Serotyping and Genotyping of GenitalChlamydia trachomatis Isolates Reveal Variants of Serovars Ba, G, and J as Confirmed by omp1 Nucleotide Sequence Analysis', *Journal of Clinical Microbiology*, 36(2), pp. 345–351. doi: 10.1128/JCM.36.2.345-351.1998.

Morré, S. A. *et al.* (2005) 'Real-time polymerase chain reaction to diagnose lymphogranuloma venereum.', *Emerging infectious diseases*, 11(8), pp. 1311–2. doi: 10.3201/eid1108.050535.

Morrison, R. P. and Caldwell, H. D. (2002) 'Immunity to murine chlamydial genital infection.', *Infection and immunity*, 70(6), pp. 2741–51. doi: 10.1128/iai.70.6.2741-2751.2002.

Morton, A. N. *et al.* (2006) 'Anorectal lymphogranuloma venereum in a Melbourne man.', *Sexual health*, 3(3), pp. 189–90. doi: 10.1071/sh06029.

Moulder, J. W. (1966) 'The Relation of the Psittacosis Group (Chlamydiae) to Bacteria and Viruses', *Annual Review of Microbiology*. Annual Reviews, 20(1), pp. 107–130. doi: 10.1146/annurev.mi.20.100166.000543.

Moulder, J. W. (1991) 'Interaction of chlamydiae and host cells in vitro.', *Microbiological reviews*, 55(1), pp. 143–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2030670.

Navarro, C. *et al.* (2002) 'Risk factors for genital chlamydial infection.', *The Canadian journal of infectious diseases = Journal canadien des maladies infectieuses*, 13(3), pp. 195–207. doi: 10.1155/2002/954837.

Nieuwenhuis, R. F. *et al.* (2003) 'Unusual presentation of early lymphogranuloma venereum in an HIV-1 infected patient: effective treatment with 1 g azithromycin.', *Sexually transmitted infections*, 79(6), pp. 453–5. doi: 10.1136/sti.79.6.453.

Nieuwenhuis, R. F. *et al.* (2004) 'Resurgence of lymphogranuloma venereum in Western Europe: an outbreak of Chlamydia trachomatis serovar l2 proctitis in The Netherlands among men who have sex with men.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 39(7), pp. 996–1003. doi: 10.1086/423966.

Nigg, C. and Eaton, M. D. (1944) 'ISOLATION FROM NORMAL MICE OF A PNEUMOTROPIC VIRUS WHICH FORMS ELEMENTARY BODIES.', *The Journal of experimental medicine*, 79(5), pp. 497–510. doi: 10.1084/jem.79.5.497.

Norrander, J., Kempe, T. and Messing, J. (1983) 'Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis', *Gene*, 26(1), pp. 101–106. doi: 10.1016/0378-1119(83)90040-9.

Nunes, A. *et al.* (2009) 'Evolutionary Dynamics of ompA, the Gene Encoding the Chlamydia trachomatis Key Antigen', *Journal of Bacteriology*, 191(23), pp. 7182–7192. doi: 10.1128/JB.00895-09.

Nwokolo, N. C. *et al.* (2016) '2015 UK national guideline for the management of infection with Chlamydia trachomatis', *International Journal of STD & AIDS*, 27(4), pp. 251–267. doi: 10.1177/0956462415615443.

O'Byrne, P. *et al.* (2016) 'Approach to lymphogranuloma venereum.', *Canadian family physician Medecin de famille canadien*, 62(7), pp. 554–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27412206.

O'Neill, C. E. *et al.* (2018) 'Genetic Transformation of a C. trachomatis Ocular Isolate With the Functional Tryptophan Synthase Operon Confers an Indole-Rescuable Phenotype.', *Frontiers in cellular and infection microbiology*, 8, p. 434. doi: 10.3389/fcimb.2018.00434.

Oakeshott, P. *et al.* (2002) 'Detection of Chlamydia trachomatis infection in early pregnancy using self-administered vaginal swabs and first pass urines: a cross-sectional community-based survey.', *The British journal of general practice : the journal of the Royal College of General Practitioners*, 52(483), pp. 830–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12392124.

Omsland, A. *et al.* (2012) 'Developmental stage-specific metabolic and transcriptional activity of Chlamydia trachomatis in an axenic medium.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(48), pp. 19781–5. doi: 10.1073/pnas.1212831109.

Oroz, C. (2001) 'Chlamydial infections in children', *Sexually Transmitted Infections*, 77(6), pp. 462–462. doi: 10.1136/sti.77.6.462.

Pannekoek, Y. *et al.* (2008) 'Multi locus sequence typing of Chlamydiales: clonal groupings within the obligate intracellular bacteria Chlamydia trachomatis.', *BMC microbiology*, 8, p. 42. doi: 10.1186/1471-2180-8-42.

Panzetta, M. E., Valdivia, R. H. and Saka, H. A. (2018) 'Chlamydia Persistence: A Survival Strategy to Evade Antimicrobial Effects in-vitro and in-vivo', *Frontiers in Microbiology*, 9. doi: 10.3389/fmicb.2018.03101.

Parra-Sánchez, M. *et al.* (2016) 'Clinical and epidemiological characterisation of lymphogranuloma venereum in southwest Spain, 2013-2015.', *Sexually transmitted infections*, 92(8), pp. 629–631. doi: 10.1136/sextrans-2015-052453.

Pedersen, L. N., Herrmann, B. and Møller, J. K. (2009) 'Typing Chlamydia trachomatis: from egg yolk to nanotechnology.', *FEMS immunology and medical microbiology*, 55(2), pp. 120–30. doi: 10.1111/j.1574-695X.2008.00526.x.

Pedersen, L. N., Pødenphant, L. and Møller, J. K. (2008) 'Highly discriminative genotyping of

Chlamydia trachomatis using omp1 and a set of variable number tandem repeats.', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 14(7), pp. 644–52. doi: 10.1111/j.1469-0691.2008.02011.x.

Peuchant, O. *et al.* (2012) 'MLVA subtyping of genovar E Chlamydia trachomatis individualizes the Swedish variant and anorectal isolates from men who have sex with men.', *PloS one*, 7(2), p. e31538. doi: 10.1371/journal.pone.0031538.

Peuchant, O. *et al.* (2016) 'Changing Pattern of Chlamydia trachomatis Strains in Lymphogranuloma Venereum Outbreak, France, 2010–2015', *Emerging Infectious Diseases*, 22(11), pp. 1945–1947. doi: 10.3201/eid2211.160247.

Public Health England (2017) *Recent trends in Lymphogranuloma Venereum (LGV) diagnoses in the United Kingdom: 2003-2016*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/653619/hpr3717_LGV.pdf.

Public Health England (2019) *Sexually transmitted infections and screening for chlamydia in England, 2018*. Available at: https://pcwhf.co.uk/wp-content/uploads/2019/06/hpr1919_stis-ncsp_ann18.pdf.

Public Health England (2020a) *LGV Data Table in England; 2011-2019*.

Public Health England (2020b) *New STI diagnoses and rates by gender, sexual risk, age group and ethnic group, 2015 to 2019*.

Public Health England (2020c) *Sexually transmitted infections and screening for chlamydia in England, 2019*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/914184/STI_NCSP_report_2019.pdf.

Public Health England (2020d) *Trends of Lymphogranuloma venereum (LGV) in England: 2019*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942751/hpr2320_LGV-10.pdf.

Public Health England (2021) *Sexually transmitted infections and screening for chlamydia in England, 2020*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1015176/STI_NCSP_report_2020.pdf.

De Puysseleyr, L. *et al.* (2017) 'Assessment of Chlamydia suis Infection in Pig Farmers.', *Transboundary and emerging diseases*, 64(3), pp. 826–833. doi: 10.1111/tbed.12446.

Qin, X. *et al.* (2016) 'Prevalence of Chlamydia trachomatis Genotypes in Men Who Have Sex with Men and Men Who Have Sex with Women Using Multilocus VNTR Analysis-ompA Typing in Guangzhou, China', *PLOS ONE*. Edited by Y.-F. Chang, 11(7), p. e0159658. doi: 10.1371/journal.pone.0159658.

Quint, K. D. *et al.* (2011) 'Anal infections with concomitant Chlamydia trachomatis genotypes among men who have sex with men in Amsterdam, the Netherlands.', *BMC infectious diseases*, 11, p. 63. doi: 10.1186/1471-2334-11-63.

Rajaram, K. et al. (2015) 'Mutational analysis of the Chlamydia muridarum plasticity zone.', *Infect Immun,* 83(7):2870-81. doi: 10.1128/IAI.00106-15.

Ramsey, K. H. et al. (2009) 'Strain and virulence diversity in the mouse pathogen Chlamydia muridarum.', *Infect Immun,* 77(8): pp. 3284-93. doi: 10.1128/IAI.00147-09.

Rasmussen, S. J., et al. (1997) 'Secretion of proinflammatory cytokines by epithelial cells in response to Chlamydia infection suggests a central role for epithelial cells in chlamydial pathogenesis.' *J Clin Invest,* 99(1):77-87. doi: 10.1172/JCI119136.

Read, T. D. *et al.* (2000) 'Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.', *Nucleic acids research*, 28(6), pp. 1397–406. doi: 10.1093/nar/28.6.1397.

Read, T. D. *et al.* (2003) 'Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae.', *Nucleic acids research*, 31(8), pp. 2134–47. doi: 10.1093/nar/gkg321.

Richardson, D. and Goldmeier, D. (2007) 'Lymphogranuloma venereum: an emerging cause of proctitis in men who have sex with men', *International Journal of STD & AIDS*, 18(1), pp. 11–16. doi: 10.1258/095646207779949916.

Robert Koch-Institut (2004) 'Zum gehauften Auftreten von Lymphogranuloma Venereum in Hamburg im Jahr 2003', *Epidemiologisches Bulletin*, 25.

Rodriguez-Dominguez, M. *et al.* (2015) 'High Prevalence of Co-Infections by Invasive and Non-Invasive Chlamydia trachomatis Genotypes during the Lymphogranuloma Venereum Outbreak in Spain', *PLOS ONE*. Edited by D. Dean, 10(5), p. e0126145. doi: 10.1371/journal.pone.0126145.

Rodríguez-Domínguez, M. *et al.* (2014) 'Clinical and epidemiological characterization of a

lymphogranuloma venereum outbreak in Madrid, Spain: co-circulation of two variants', *Clinical Microbiology and Infection*, 20(3), pp. 219–225. doi: 10.1111/1469-0691.12256.

Rodríguez-Marañón, M. J. *et al.* (2002) 'Prediction of the membrane-spanning beta-strands of the major outer membrane protein of Chlamydia.', *Protein science : a publication of the Protein Society*, 11(7), pp. 1854–61. doi: 10.1110/ps.3650102.

Rönn, M. *et al.* (2014) 'Characteristics of LGV repeaters: analysis of LGV surveillance data.', *Sexually transmitted infections*, 90(4), pp. 275–8. doi: 10.1136/sextrans-2013-051386.

Sachse, K. *et al.* (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.', *Systematic and applied microbiology*, 37(2), pp. 79–88. doi: 10.1016/j.syapm.2013.12.004.

Sachse, K. *et al.* (2015) 'Emendation of the family Chlamydiaceae: proposal of a single genus, Chlamydia, to include all currently recognized species.', *Systematic and applied microbiology*, 38(2), pp. 99–103. doi: 10.1016/j.syapm.2014.12.004.

Sanders, A. M. *et al.* (2019) 'Prevalence of trachoma within refugee camps serving South Sudanese refugees in White Nile State, Sudan: Results from population-based surveys', *PLOS Neglected Tropical Diseases*. Edited by R. Reithinger, 13(6), p. e0007491. doi: 10.1371/journal.pntd.0007491.

Satoh, M. *et al.* (2014) 'Multilocus VNTR analysis-ompA typing of venereal isolates of Chlamydia trachomatis in Japan.', *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 20(10), pp. 656–9. doi: 10.1016/j.jiac.2014.06.010.

Savage, E. J. *et al.* (2014) 'Improving surveillance of sexually transmitted infections using mandatory electronic clinical reporting: the genitourinary medicine clinic activity dataset, England, 2009 to 2013.', *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 19(48), p. 20981. doi: 10.2807/1560-7917.es2014.19.48.20981.

Saxon, C. *et al.* (2016) 'Asymptomatic Lymphogranuloma Venereum in Men who Have Sex with Men, United Kingdom.', *Emerging infectious diseases*, 22(1), pp. 112–116. doi: 10.3201/EID2201.141867.

Schachter, J. *et al.* (2001) 'Radical changes to chlamydial taxonomy are not necessary just yet.', *International journal of systematic and evolutionary microbiology*, 51(Pt 1), p. 249; author reply 251-3. doi: 10.1099/00207713-51-1-249.

Scherchand, S. and Aiyar, A. (2019) 'Ammonia generation by tryptophan synthase drives a key genetic difference between genital and ocular Chlamydia trachomatis isolates.', *Proc Natl Acad Sci USA,* 116(25): pp. 12468-12477. doi: 10.1073/pnas.1821652116.

Schoborg, R. V. (2011) 'Chlamydia persistence – a tool to dissect chlamydia–host interactions', *Microbes and Infection*, 13(7), pp. 649–662. doi: 10.1016/j.micinf.2011.03.004.

Scieux, C. *et al.* (1989) 'Lymphogranuloma Venereum: 27 Cases in Paris', *Journal of Infectious Diseases*, 160(4), pp. 662–668. doi: 10.1093/infdis/160.4.662.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–9.

Seemann, T. (2018) *Snippy*. Available at: https://github.com/tseemann/snippy.

Seth-Smith, H. M. B. *et al.* (2013) 'Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture', *Genome Research*, 23(5), pp. 855–866. doi: 10.1101/gr.150037.112.

Seth-Smith, H. M. B. *et al.* (2017) 'European Chlamydia abortus livestock isolate genomes reveal unusual stability and limited diversity, reflected in geographical signatures.', *BMC genomics*, 18(1), p. 344. doi: 10.1186/s12864-017-3657-y.

Seth-Smith, H. M. B. *et al.* (2021) 'Ongoing evolution of Chlamydia trachomatis lymphogranuloma venereum: exploring the genomic diversity of circulating strains.', *Microbial genomics*, 7(6). doi: 10.1099/mgen.0.000599.

Shankar, M. *et al.* (2006) 'Prevalence of Chlamydia trachomatis and bacterial vaginosis in women presenting to the early pregnancy unit.', *Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology*, 26(1), pp. 15–9. doi: 10.1080/01443610500363899.

da Silva, M. A. L. *et al.* (2013) 'Optimization of single-tube nested PCR for the diagnosis of visceral leishmaniasis.', *Experimental parasitology*, 134(2), pp. 206–10. doi: 10.1016/j.exppara.2013.03.003.

Skilton, R. J. et al. (2009) 'Penicillin induced persistence in Chlamydia trachomatis: high quality time lapse video analysis of the developmental cycle.' *PLoS One,* 4(11):e7723. doi: 10.1371/journal.pone.0007723.

Skipp, P. J. S. *et al.* (2016) 'Quantitative Proteomics of the Infectious and Replicative Forms of Chlamydia trachomatis.', *PloS one*, 11(2), p. e0149011. doi: 10.1371/journal.pone.0149011.

Somboonna, N. *et al.* (2011) 'Hypervirulent Chlamydia trachomatis Clinical Strain Is a Recombinant between Lymphogranuloma Venereum (L2) and D Lineages', *mBio*. Edited by L. A. Campbell and S. Maloy, 2(3). doi: 10.1128/mBio.00045-11.

Soni, S. *et al.* (2010) 'Lymphogranuloma venereum proctitis masquerading as inflammatory bowel disease in 12 homosexual men', *Alimentary Pharmacology & Therapeutics*, 32(1), pp. 59–65. doi: 10.1111/j.1365-2036.2010.04313.x.

Spaargaren, J., Fennema, H. S. A., *et al.* (2005) 'New lymphogranuloma venereum Chlamydia trachomatis variant, Amsterdam.', *Emerging infectious diseases*, 11(7), pp. 1090–2. doi: 10.3201/eid1107.040883.

Spaargaren, J., Schachter, J., *et al.* (2005) 'Slow Epidemic of Lymphogranuloma Venereum L2b Strain', *Emerging Infectious Diseases*, 11(11), pp. 1787–1788. doi: 10.3201/eid1111.050821.

Stamm, W. E. (1999) 'Chlamydia trachomatis infections: progress and problems.', *The Journal of infectious diseases*, 179 Suppl, pp. S380-3. doi: 10.1086/513844.

Stary, G. *et al.* (2008) 'New Chlamydia trachomatis L2 strains identified in a recent outbreak of lymphogranuloma venereum in Vienna, Austria.', *Sexually transmitted diseases*, 35(4), pp. 377–82. doi: 10.1097/OLQ.0b013e31815d6df8.

Stephens, R. S. *et al.* (1982) 'Monoclonal antibodies to Chlamydia trachomatis: antibody specificities and antigen characterization.', *Journal of immunology (Baltimore, Md. : 1950)*, 128(3), pp. 1083–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7035557.

Stephens, R. S. *et al.* (1987) 'Diversity of Chlamydia trachomatis major outer membrane protein genes.', *Journal of bacteriology*, 169(9), pp. 3879–85. doi: 10.1128/jb.169.9.3879-3885.1987.

Stephens, R. S. *et al.* (2001) 'Heparin-binding outer membrane protein of chlamydiae.', *Molecular microbiology*, 40(3), pp. 691–9. doi: 10.1046/j.1365-2958.2001.02418.x.

Stephens, R. S. *et al.* (2009) 'Divergence without difference: phylogenetics and taxonomy of Chlamydia resolved.', *FEMS immunology and medical microbiology*, 55(2), pp. 115–9. doi: 10.1111/j.1574-695X.2008.00516.x.

Storz, J. and Page, L. A. (1971) 'Taxonomy of the Chlamydiae: Reasons for Classifying Organisms of the Genus Chlamydia, Family Chlamydiaceae, in a Separate Order, Chlamydiales ord. nov.', *International Journal of Systematic Bacteriology*, 21(4), pp. 332–334. doi: 10.1099/00207713-21-4-332.

Su, H. *et al.* (1990) 'Chlamydia trachomatis-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin.', *Infection and immunity*, 58(4), pp. 1017–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2318528.

Su, H. *et al.* (1996) 'A recombinant Chlamydia trachomatis major outer membrane protein binds to heparan sulfate receptors on epithelial cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), pp. 11143–8. doi: 10.1073/pnas.93.20.11143.

Swanson, J. *et al.* (1975) 'Light and Electron Microscopic Study of Chlamydia trachomatis Infection of the Uterine Cervix', *Journal of Infectious Diseases*, 131(6), pp. 678–687. doi: 10.1093/infdis/131.6.678.

Sykes, J. E. (2005) 'Feline chlamydiosis.', *Clinical techniques in small animal practice*, 20(2), pp. 129–34. doi: 10.1053/j.ctsap.2004.12.018.

Szabo, K. V., O'Neill, C. E. and Clarke, I. N. (2020) 'Diversity in Chlamydial plasmids', *PLOS ONE*. Edited by D. M. Ojcius, 15(5), p. e0233298. doi: 10.1371/journal.pone.0233298.

Thompson, C. (2001) 'Lesson of the week: A family cluster of Chlamydia trachomatis infection', *BMJ*, 322(7300), pp. 1473–1474. doi: 10.1136/bmj.322.7300.1473.

Thomson, N. R. *et al.* (2007) 'Chlamydia trachomatis: Genome sequence analysis of lymphogranuloma venereum isolates', *Genome Research*, 18(1), pp. 161–171. doi: 10.1101/gr.7020108.

Touati, A. *et al.* (2016) 'The L2b real-time PCR targeting the pmpH gene of Chlamydia trachomatis used for the diagnosis of lymphogranuloma venereum is not specific to L2b strains.', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 22(6), p. 574.e7-9. doi: 10.1016/j.cmi.2016.03.018.

Traeger, M. W. *et al.* (2018) 'Effects of Pre-exposure Prophylaxis for the Prevention of Human Immunodeficiency Virus Infection on Sexual Risk Behavior in Men Who Have Sex With Men: A Systematic Review and Meta-analysis', *Clinical Infectious Diseases*, 67(5), pp. 676–686. doi: 10.1093/cid/ciy182.

Traeger, M. W. *et al.* (2019) 'Association of HIV Preexposure Prophylaxis With Incidence of Sexually Transmitted Infections Among Individuals at High Risk of HIV Infection.', *JAMA*, 321(14), pp. 1380–1390. doi: 10.1001/jama.2019.2947.

United Nations High Commissioner for (2019) *UNHCR Global Trends - Forced Displacement in 2017*. Available at: https://www.unhcr.org/globaltrends2017/.

Urwin, R. and Maiden, M. C. J. (2003) ‘Multi-locus sequence typing: a tool for global epidemiology’, *Trends in Microbiology*, 11(10), pp. 479–487. doi: 10.1016/j.tim.2003.08.006.

Vandenbruaene, M. (2004) ‘Uitbraak van Lymphogranuloma Venereum in Antwepen en Rotterdam’, *Epidemiologisch Bulletin van de Vlaamse Gemeenschap*, 25.

Versteeg, B., et al. (2020) 'Viability PCR shows that non-ocular surfaces could contribute to transmission of Chlamydia trachomatis infection in trachoma.', *PLoS Negl Trop Dis,* 14(7):e0008449. doi: 10.1371/journal.pntd.0008449.

Vorimore, F. *et al.* (2013) ‘Isolation of a New Chlamydia species from the Feral Sacred Ibis (Threskiornis aethiopicus): Chlamydia ibidis.’, *PloS one*, 8(9), p. e74823. doi: 10.1371/journal.pone.0074823.

de Vries, H. *et al.* (2019) *2019 European Guidelines on the Management of Lymphogranuloma Venereum*.

de Vries, H. J. C. *et al.* (2019) ‘2019 European guideline on the management of lymphogranuloma venereum’, *Journal of the European Academy of Dermatology and Venereology*, 33(10), pp. 1821–1828. doi: 10.1111/jdv.15729.

de Vrieze, N. H. N. *et al.* (2017) ‘Low Prevalence of Urethral Lymphogranuloma Venereum Infections Among Men Who Have Sex With Men’, *Sexually Transmitted Diseases*, 44(9), pp. 547–550. doi: 10.1097/OLQ.0000000000000657.

Wang, S. P. and Grayston, J. T. (1970) ‘Immunologic relationship between genital TRIC, lymphogranuloma venereum, and related organisms in a new microtiter indirect immunofluorescence test’, *Am J Ophthalmol*, 70, pp. 367–374.

Wang, Y. *et al.* (2011) ‘Evaluation of a high resolution genotyping method for Chlamydia trachomatis using routine clinical samples.’, *PloS one*, 6(2), p. e16971. doi: 10.1371/journal.pone.0016971.

Ward, H. *et al.* (2007) ‘Lymphogranuloma venereum in the United kingdom.’, *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 44(1), pp. 26–32. doi: 10.1086/509922.

Wayal, S. *et al.* (2017) ‘Ethnic variations in sexual behaviours and sexual health markers: findings from the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3).’, *The Lancet. Public health*, 2(10), pp. e458–e472. doi: 10.1016/S2468-2667(17)30159-7.

Weir, E., Haider, S. and Telio, D. (2004) 'Trachoma: leading cause of infectious blindness.', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 170(8), p. 1225. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15078842.

Willcox, M. D. P., (2013) 'Characterisation of the normal microbiota of the ocular surface.' *Exp Eye Res,* 117:99-105. doi: 10.1016/j.exer.2013.06.003.

World Health Organisation (1987) *Trachoma Simplified Grading Card*. Available at: http://www.who.int/trachoma/resources/SAFE_documents/en/ (Accessed: 31 July 2019).

World Health Organisation (2020a) *A road-map for neglected tropical diseases: 2021-2030*. Available at: https://www.who.int/neglected_diseases/Ending-the-neglect-to-attain-the-SDGs--NTD-Roadmap.pdf?ua=1.

World Health Organisation (2020b) *Trachoma Factsheet*. Available at: https://www.who.int/news-room/fact-sheets/detail/trachoma.

World Health Organization (1997) *Report of the First Meeting of the Who alliance for the Global Elimination of Trachoma*, *World Health Organization*. Available at: http://www.afro.who.int/sites/default/files/2017-06/1997_0.pdf.

World Health Organization (2016) *Global Health Sector Strategy on Sexually Transmitted Infections 2016-2021*. Available at: https://apps.who.int/iris/bitstream/handle/10665/246296/WHO-RHR-16.09-eng.pdf;sequence=1.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) 'Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.', *Gene*, 33(1), pp. 103–19. doi: 10.1016/0378-1119(85)90120-9.

Yao, J. and Rock, C. O. (2018) 'Therapeutic Targets in Chlamydial Fatty Acid and Phospholipid Synthesis', *Frontiers in Microbiology*, 9. doi: 10.3389/fmicb.2018.02291.

Zar, H. J. (2005a) 'Neonatal chlamydial infections: prevention and treatment.', *Paediatric drugs*, 7(2), pp. 103–10. doi: 10.2165/00148581-200507020-00003.

Zar, H. J. (2005b) 'Neonatal Chlamydial Infections', *Pediatric Drugs*, 7(2), pp. 103–110. doi: 10.2165/00148581-200507020-00003.