UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE,HEALTH AND LIFE SCIENCES

School of Medicine

The Molecular Epidemiology of *Chlamydia trachomatis* [Genotypes In Comparison With Disease Presentations In Patients In Jersey]

By

Sarah Jane Lockey BSc

A thesis submitted for the degree of Master of Philosophy

December 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE,HEALTH AND LIFE SCIENCES

School of Medicine

Master of Philosophy

THE MOLECULAR EPIDEMIOLOGY OF *CHLAMYDIA TRACHOMATIS* [GENOTYPES IN COMPARISON WITH DISEASE PRESENTATIONS IN PATIENTS IN JERSEY]

By Sarah Jane Lockey

The Chlamydiaceae are a family of small obligate intracellular bacteria, parasitic to eukaryotic cells that replicate inside a vacuole known as an inclusion. *Chlamydia trachomatis* is the causative agent of trachoma, inclusion conjunctivitis, and Sexually Transmitted Infections including Lymphogranuloma venereum.

C.trachomatis positive clinical samples were identified via the Becton Dickinson (BD) Probe Tech method in Jersey and via the Roche Cobas Taqman method in Southampton. Details of all samples were anonymised. Amplification of clinical samples template DNA for visualisation was completed by PCR. All sample batches received from Jersey were screened for *C.trachomatis* positivity using *ompA* specific primers (Ngandjio *et al.,* 2003) (Jurstrand *et al.,* 2001). Positive samples were amplified to give product sufficient for gel extraction and automated DNA sequencing to confirm serovar genotyping.

A new variant of *C.trachomatis* with a deletion in the cryptic plasmid has been detected in Sweden, following a 25% decrease in *C.trachomatis* infections between November 2005 and August 2006 in Halland county. Part of the cryptic plasmid has been already sequenced from the variant strain. There is a deletion of 377 base pairs in the target area for the Abbott and Roche *C.trachomatis* NAAT tests. This deletion does not affect Beckton Dickinson *C.trachomatis* test. The cryptic plasmid deletion was detected using specific primers that had been mapped either side of the deletion (Ripa and Nilsson, 2006).

TABLE OF CONTENTS

TITLE	**I**
ABSTRACT	**II**
LIST OF CONTENTS	**III**
LIST OF FIGURES	**VI**
LIST OF TABLES	**VIII**
PUBLICATIONS	**IX**
AUTHOR'S DECLARATION	**IX**
ACKNOWLEDGMENTS	**X**
ABBREVIATIONS USED	**XI**

CHAPTER 1 INTRODUCTION	**1**

1.0	Chlamydia	1
1.1	Chlamydia historical background	2
1.2	Chlamydia taxonomy and classification	8
1.3	Biological differentiation within *C.trachomatis*	11
1.4	Epidemiology of *Chlamydia trachomatis*	14
1.4.1	Review of the Epidemiology in Europe	17
1.4.2	The Netherlands	17
1.4.3	Sweden	20
1.4.4	Denmark	22
1.4.5	United Kingdom	24
1.5	Project aims	27
1.5.1	Primary objectives of the study	28

CHAPTER 2 MATERIALS AND METHODS		

2.0	Culture of Chlamydiae	30
2.0.1	Bacterial strains	30
2.1	Media	31
2.2	Buffers, reagents and solutions	33
2.3	UHQ Water	33
2.4	Chlamydia host cell culture	34
2.4.1	Chlamydia host cell storage	34
2.4.2	Chlamydia host cell resuscitation	34
2.4.3	Chlamydia host cell passage	35
2.5	Chlamydia culture	36
2.5.1	LGV biovar strains	37
2.5.2	Trachoma biovar strains	37
2.6	Chlamydia harvesting with glass beads	38
2.6.1	Chlamydia harvesting with glass beads for DNA	38
2.7	Chlamydia harvesting with a homogeniser	39
2.8	Separation of EBs and RBs with Urografin	40
2.8.1	First gradient	40
2.8.2	Second gradient	41
2.9	Chlamydia positive clinical samples	42
2.10	Genetic analysis	43
2.10.1	DNA Extraction methods	43
2.10.1.1	Extraction of chlamydial genomic DNA from laboratory cultivated stains	43
2.10.1.2	Extraction of chlamydial genomic DNA from clinical samples via ion exchange gel purification	44
2.10.1.3	Extraction of chlamydial genomic DNA from clinical samples via MagNA Pure LC total nucleic acid isolation kit (Roche)	44

2.10.1.4	Extraction of chlamydial genomic DNA from clinical samples via QIAamp spin columns (Qiagen)	46
2.11	Quantitation of DNA by agarose gel electrophoresis	46
2.12	Primer design and selection	47
2.13	Amplification of chlamydial DNA by Polymerase Chain Reaction (PCR)	48
2.13.1	PCR reaction components	49
2.13.2	Thermocycling conditions	50
2.14	PCR amplicon detection, characterisation and quantification via agarose gel electrophoresis	51
2.14.1	Visualisation of DNA	53
2.14.2	Estimation of DNA size and concentration	53
2.15	1 kb ladder	53
2.16	1 kb plus ladder	54
2.17	Purification of amplified products	54
2.17.1	Gene-Elute gel extraction kit (Sigma)	55
2.17.2	Qiaquick gel extraction kit (Qiagen)	55
2.18	Automated DNA sequencing	56
2.18.1	Materials	58
2.18.2	Sequencing reactions	58
2.18.3	Gel preparation	59
2.18.4	Gel electrophoresis	59
2.19	Computer analysis	60

CHAPTER 3: PCR DEVELOPMENT

3.0	Samples available for this study from Jersey	61
3.1	Development of *ompA* PCR using template DNA from BD Probe Tech prepared samples	63
3.1.1	Laboratory cultivated LGV strains	68
3.1.2	L2/Amsterdam/Protitis DNA concentration estimation	69
3.1.3	L2/434/Bu/I DNA concentration estimation	70
3.1.4	CT1 and CT5 primer to template specificity	71
3.2	CT1 and CT5 primer optimisation	72
3.2.1	Thermostable DNA polymerase optimisation	73
3.2.2	$MgCl_2$ concentration optimisation	75
3.3	BD Probe Tech samples selected to trial optimised PCR conditions	76
3.4	Initial *ompA* screening of ten clinical samples from Jersey	78
3.5	Dilution of original samples to potentially dilute PCR inhibitors	80
3.6	Adjustments to the reaction volume to potentially reduce the effects of PCR inhibitors without altering the amount of DNA template entering the reaction	82
3.7	Comparative evaluation of primer annealing temperature	86
3.8	Amplification of specific *ompA* variable domains	88
3.9	DNA template clean up methods	93
3.10	Chlamydial DNA template purification	94
3.11	DNA sequencing of ompA positive samples after clean up of template DNA entering the PCR reaction	96
3.12	Pooling of identical PCR products to give a more concentrated PCR product template for DNA sequencing	96
3.13	Comparisons in agarose gel staining methods	98

CHAPTER 4: SEQUENCING DEVELOPMENT

4.0	Sequencing development	102
4.1	Comparative analysis of *ompA* sequences	103
4.1.1	Comparative analysis Of LGV2 and sub strains	103
4.2	Comparative analysis of Nills Lyke and IOL1883 strains	105
4.3	Screening of samples received from Jersey	107

CHAPTER 5: COMPARATIVE ANALYSIS OF DIAGNOSTIC PLATFORMS

5.0	*OmpA* screening of Jersey and Southampton samples	122
5.1	New C.trachomatis Swedish variant deletion in the cryptic plasmid	126

CHAPTER 6: NEW VARIANT *C.TRACHOMATIS* FROM SWEDEN

6.0	*OmpA* genotyping of the nvCt	129
6.1	Confirmation of the plasmid deletion	131

CHAPTER 7: MOLECULAR GENOTYPING OF PLASMIDS

7.0	Comparative analysis of plasmids	135
7.1	Plasmid sequence selection	135
7.2	Quality of the aligned sequences and further selection	136
7.3	Intergenic analysis from the origin of replication	137
7.4	ORF 1	138
7.5	ORF 2	141
7.6	ORF 3	143
7.7	ORF 4	145
7.8	ORF 5	147
7.9	ORF 6	151

7.10	ORF 7	151
7.11	ORF 8	152
7.12	Intergenic point mutations	154
7.13	Plasmid sequence analysis of unique features	154

CHAPTER 8: GENERAL DISCUSSION AND FUTURE WORK

8.0	Jersey *ompA* project	156

REFERENCES	**161**
APPENDICES	

TABLE OF FIGURES

Figure 3.1	CT1 and CT5 primer location and PCR product size (Ngandjio *et al.,* 2003)	67
Figure 3.2	L2/Amsterdam/Proctitis original genomic DNA	69
Figure 3.3	L2/434/Bu/I original genomic DNA	70
Figure 3.4	Primer – to – template specificity testing, detection of *ompA* in DNA from two L2 isolates usingCT1 and CT5 primers	71
Figure 3.5	Evaluation of thermostable DNA Polymerases (Invitrogen and Bio-X-Act)	73
Figure 3.6	Evaluation of thermostable DNA polymerases (Promega and Red Taq)	74
Figure 3.7	Optimisation of the *ompA* specific PCR $MgCl_2$ concentration using CT1 and CT5 primers and L2/Amsterdam/Proctitis (1.6ng/µL) as template	75
Figure 3.8	Six selected samples chosen to trial optimised *ompA* specific PCR amplification conditions	77
Figure 3.9	Initial *ompA* Screen of the first 10 clinical samples received from Jersey	79
Figure 3.10	Dilution of template DNA entering the *ompA* PCR	81
Figure 3.11	Comparative evaluation of reaction volume	83
Figure 3.12	Comparative evaluation of reaction volume	84
Figure 3.13	Comparative evaluation of reaction volume	85
Figure 3.14	Comparative evaluation of primer annealing temperature	87
Figure 3.15	Primer location and PCR product size	89
Figure 3.16	Comparative evaluation of primer combinations outlined in the methods described by Jurstrand *et al.,* 2001 and Ngandjio *et al.,* 2003	90
Figure 3.17	Comparative evaluation of primer combinations outlined in the methods described by Jurstrand *et al.,* 2001and Ngandjio *et al.,* 2003	91

Figure 3.18	Recovery gel to demonstrate the gel extraction and pooling of PCR reactions from samples received from Jersey	97
Figure 3.19	Comparative evaluation of staining methods ethidium bromide	98
Figure 3.20	*OmpA* specific PCR amplicons stained with SYBR green	100
Figure 4.1	Comparative alignment of variable domain IV, serovar D	106
Figure 4.2	A comparative alignment of variable domain IV sequence data from 2005 samples	109
Figure 4.3	A comparative alignment of variable domain IV sequence data from 2006 samples	114
Figure 4.4	A comparative alignment of variable domain IV sequence data from 2007 samples	119
Figure 6.1	Sequence alignment showing the genotype E determinants of variable domain IV from a published serovar E and the swedish variant	130
Figure 6.2	Location of the new variant 377bp deletion	132
Figure 6.3	Optimisation of $MgCl_2$ for the detection of the swedish variant 377bp deletion in the cryptic plasmid	133

TABLE OF TABLES

2.1	Chlamydia trachomatis strains sourced from ATCC	30
2.2	Chlamydia trachomatis strains kindly provided by collaborators	30
2.3	Contents of DMEM (Invitrogen)	31
2.4	Buffers, reagents and solutions	33
2.5	Chlamydial host cell lines	34
2.6	20% Urografin gradient	40
2.7	34%, 44% and 54% Urografin gradient	41
2.8	MagNA Pure reagent details	45
2.9	Primer details for *ompA* specific amplification (Ngandjio *et al.,* 2003)	47
2.10	Primer details for *ompA* specific amplification (Jurstrand *et al.,*2001)	47
2.11	Primer details for *ompA* variable domain specific amplification (Ngandjio *et al.,* 2003)	48
2.12	Guide-lines for agarose gel concentrations	51
2.13	Gel dimension specifications	52
2.14	DNA ladder band sizes 1 kb	53
2.15	DNA ladder band sizes 1 kb plus	54
2.16	Amount of template DNA required per sequencing reaction	58
3.1	2005 annual prevalence of *C.trachomatis* on Jersey	61
3.2	BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2005	62
3.3	BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2006	62
3.4	BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2007	63
3.5	Details of the six selected samples chosen to trial the optimised *ompA* specific PCR amplification conditions	77

3.6	Details of five samples received from Jersey selected to test three DNA template clean up methods	94
3.7	Detection of *ompA* DNA from three DNA template clean up methods using CT1 and CT5 primers	95
4.1	Sequencing results of VDIV genotyped samples received from Jersey during 2005	107
4.2	Sequencing results of VIV genotyped samples received from Jersey during 2006	113
4.3	Sequencing results of VDIV genotyped samples received from Jersey during 2007	118
7.1	Plasmid sequence details for the alignment	136
7.2	Changes in ORF 1 that do not result in a change to the translated sequence (non-synonymous mutations)	138
7.3	Changes in ORF 2 that do not result in a change to the translated sequence (non-synonymous mutations)	141
7.4	Changes in ORF 3 that do not result in a change to the translated sequence (non-synonymous mutations)	143
7.5	Changes in ORF 4 that do not result in a change to the translated sequence (non-synonymous mutations)	145
7.6	Changes in ORF 5 that do not result in a change to the translated sequence (non-synonymous mutations)	147
7.7	Changes in ORF 8 that do not result in a change to the translated sequence (non-synonymous mutations)	153
7.8	Examples of unique features in the ORF 1 region of the plasmid compared to CTB genotype	155
8.1	TARP and PMP primer sequences	159

PUBLICATION

The work based in this thesis has contributed to the following publication

Thomson NR, Holden MT, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, Harris B, Ormond D, Rance R, Quail MA, Parkhill J, Stephens RS, Clarke IN. *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res. 2008 Jan;18(1):161-71. Epub 2007 Nov 21

ACKNOWLEDGMENTS

This work was supported by HOPE (The Wessex Medical Trust) and the University of Southampton, The Southampton HPA and the Pathology laboratory of St Helier General Hospital, Jersey. I am extremely grateful to my supervisors Prof. Ian Clarke and Dr Peter Marsh for the supervision and guidance during the execution and writing of this thesis.

Thanks are also due to Dr Sue O'Connell, from Southampton Health Protection Agency and Dr Ivan Muscat from the Pathology laboratory of St Helier General Hospital, Jersey.

I am especially indebted to the many staff from the health protection agency molecular microbiology and serology labs at Southampton General Hospital and to Alison Cabot and Stella Bowgen of the Pathology laboratory of St Helier General Hospital, Jersey. who gave their time to help collect screened *C.trachomatis* positive samples without which this research would not have been possible.

I would like to thank Dr Nick Thomson from the Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute for his invaluable help in understanding bacterial genetics and the ACT and Artemis computer programs used during this project. I am grateful to Dr Helena Seth-Smith and Ms. Nicola Lennard for independently verifying the suitability of my designed oligos and subsequently sequencing and aligning the products.

I am grateful to Mrs Lesley Cutcliffe for sharing her expertise in chlamydial culture and technical advice, Miss Rachel Skilton for sharing her technical knowledge during this project, Dr Steve Green for the contribution of his extensive expertise of PCR, Mrs Jenny Russell for her continued support throughout this project both as a lab manager and for patiently listening to my 'rantings' and redirecting my frustrations into constructive progress, and to Karen and Shirley for ensuring all consumables were to hand, a very big thank you.

ABBREVIATIONS

BD	Beckton Dickinson
BGMK	Buffalo Green Monkey Kidney
bp	Base Pair
ddNTP	Dideoxyribonucleoside triphosphate
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside Triphosphate
EB	Elementary Body
EDTA	Ethylenediaminetetraacetic Acid
FRET	Fluorescence Resonance Energy Transfer
LGV	Lymphogranuloma vanereum
MOMP	Major Outer Membrane Protein
MOTA	Method Other Than Acceleration
NAAT	Nucleic Acid Amplification Test
nvCt	New Variant Chlamydia trachomatis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RB	Reticulate Body
SDA	Strand Displacement Amplification
Taq	Thermostable polymerase
TE	Tris/EDTA
TEMED	Tetramethylethylenediamine
UHQ	Ultra High Quality
2SP	2 x Sucrose Phosphate
4SP	4 x Sucrose Phosphate

CHAPTER 1 INTRODUCTION

1.0 Chlamydia

Members of the Chlamydiaceae family have been identified as the causative agents of a wide range of human and animal based diseases. Chlamydiae are small gram negative bacteria that are able to synthesize Adenosine Tri-Phosphate (ATP) and Guanine Tri-Phosphate (GTP) however, they remain obligate intracellular pathogens.

Chlamydiae have an alternating two stage developmental cycle. During the first stage the chlamydiae are infectious and known as Elementary Bodies (EBs). EBs are ~300nm in diameter (Schachter, 1990) and are specialised for extra-cellular survival. EBs bind to susceptible host cells and subsequently infect them, in doing this they are taken up into an intracellular vacuole known as an inclusion. This inclusion is used by the chlamydiae to transform into their second stage the Reticulate Body (RBs). RBs are non infectious and are only present within inclusions, the RBs are the intracellular replicating form of the micro-organism (Schachter, 1990).

1.1 Chlamydia historical background

Chlamydia are known to have been infectious to humans for millennia. The earliest surviving document, The Ebers papyrus c.1500BC, refers to what was much later classified as *Chlamydia* as 'an exudative cicatrising eye disease'. The ancient Greeks also knew of *Chlamydia,* as did the Chinese and the Romans. Notable sufferers from that historical period included Paul of Tarsus, Cicero, Horace and Pliny the Younger. In 60AD a Sicilian physician, Pedanius Diascarides, coined the name 'trachoma' meaning rough eye for the disease then presenting in his clinic (Oriel & Ridgway, 1982).

In 1786 John Hunter first described Lympho-Granuloma Venereum (LGV). Nearly a century later, 1879, Neisser discovered the gonococcus and soon afterwards the Gram staining method was described. Gram staining showed the difference between gonococcal forms and non-gonococcal forms of urethritis and ophthalmia neonatorum. In 1884 Kroner suggested that an unknown infective agent in the mother's genital tract might cause non-gonococcal ophthalmia neonatorum. However, the venereal link between the urethritis in males and the symptoms in females and babies remained unrecognised (Oriel & Ridgway, 1982).

In 1883, Koch carried out the first study of trachoma aetiology, although it was his intention to investigate Egyptian Ophthalmia. During this study Koch recovered *Haemophilus aegyptius* and *Neisseria gonorrhoeae* and concluded that the latter disease was synonymous with trachoma.

However, it was not until 1907, when the work of Halberstaedter and Von Prowazek was published, that it was inferred that the infective agent of trachoma was not *N.gonorrhoeae.* Halberstaedter and Von Prowazek described trachoma and the intraepithelial inclusion bodies that trachoma forms. Conjunctival scrapings were taken from experimentally infected Orang-Utans,

and stained with giemsa. The scrapings showed intracellular vacuoles, inside which the Chlamydiae grew. The paper described these chlamydial inclusions as 'intracytoplasic vacuoles containing numerous minute particles'. In later work Halberstaedter and Von Prowazek, (1909), described identical inclusion bodies in the conjunctival smears of babies infected with non-gonococcal ophthalmia neonatorum.

The link between non gonococcal urethritis (NGU), the birth canal and the conjunctival infection in babies remained unproven. In 1910 Hayman recorded the presence of identical inclusion bodies in the urethral and cervical cells of parents who had babies who presented with non-gonococcal ophthalmia neonatorum. Unfortunately, this work was not able to draw completely conclusive evidence for the chlamydial link between the birth canal, male urethritis and ophthalmia neonatorum as some of the subjects had mixed gonococcal and non-gonococcal infections.

The lack of conclusive evidence was short lived as Fritsch *et al.*, (1910) (Oriel, J. D. & Ridgway, G. L. (1982)) discovered a firm link between the ocular and genital infections of *C.trachomatis*. The research compared monkey conjunctivae infected with; conjuctival scrapings from babies or ophthalmia neonatorum, or cervical secretions of their mothers or urethral secretions of males with NGU. All the infected monkeys presented with inclusion conjunctivitis that was identical in pathology irrespective of the source of infection.

A year later Von Wahl (1911) stated post gonococcal urethritis (PGU) infections to be as a result of mixed infections of *N.gonorrhoeae* and the infectious agent of NGU, now known to be *C.trachomatis.* Advances in *C.trachomatis* study were constrained by the fact that the organism could not be cultured *in vitro*. Some ocular and genital infections were cultured in the

conjunctivae of blind volunteers for confirmation of *C.trachomatis* during the 1930's and lead to similar studies completed by Thygeson and others (Oriel & Ridgway, 1982; Thygeson & Stone, 1942; Thygeson & Mengert, 1936). Finally in 1932, the organism responsible for LGV was first isolated from cerebral inclusions in monkeys, and a year later LGV strains were successfully cultivated in mice.

There was a huge outbreak of psittacosis among birds in 1929. Fears that the agent would transmit to the human population sparked a renewed wave of interest into the Chlamydiaceae family. Bedson and Bland, 1934 clearly described the developmental cycle for the psitticosis agent, which had been previously isolated from birds, and likened it to the developmental cycle of *C.trachomatis*.

Treatment of *C.trachomatis* infections with sulfonamides was commonly practiced by General Practitioners in the 1930's even though there was an inability to grow the chlamydial organism in a laboratory. The pharmacological mode of action would remain unknown until the work by Kurnosova and Lenkevich, 1964.

In an epiderniological study of chlamydial infection carried out in 1942 by Thygeson and Stone, one hundred men with urethritis were examined. Seven of the examined men presented with mixed gonococcal and chlamydial infections and two had solely chlamydial infections. The type of infection was confirmed by inoculation of baboon conjunctivae with swabs taken from the participants. The paper noted the high incidence of babies contracting inclusion conjunctivitis that had fathers with a history of urethritis, whether the participants of the study with solely gonococcal infections developed PGU at a later date was not recorded.

Despite all the work published before 1945 there was still some doubt about PGU. This was demonstrated in 1948 when Mast quoted, by Harkness, 1950, claimed there was a new disease caused by penicillin-resistant gonorrhoea on a US Navy ship. Harkness noted that inclusions were present in most incidences of NGU but not consistently so. In addition scientists were unable to culture the organism in the laboratory using filtrates of the non-gonococcal discharges on the chorio-allantoic membrane of fertilised hen's eggs.

T'ang *et al.,* (1957) successfully isolated the trachoma agent using the methods of Cox, (1938) who described the technique to isolate Rickettsiae using the yolk membrane of fertilised hen's eggs. The eggs were initially treated with penicillin to stop them from presenting with super infections. But pre-treatment was soon found to adversely affect the isolation of the chlamydiae. Tetracycline sensitivity and resistance to aminoglycosides by *C.trachomatis* was also noted. The success of the new laboratory process to culture the organism led to a rapid revival of interest in chlamydiae.

Collier and Sowa, (1958), demonstrated the serological relationship between the yolk sac isolate and members of the trachoma group. The availability of the culture technique in eggs also enabled laboratory researchers to re-examine the role of *C.trachomatis* in the genital tract.

Collier *et al.,* (1958) also established that *C.trachomatis,* the agent of trachoma, was transmitted from human to human. The establishment of human to human transmission led to significant research into the prevention and control of trachoma which included an attempt to produce a vaccine. Unfortunately, to date vaccination has been unsuccessful, and is still the goal of much chlamydial research.

Jones *et al.,* (1959) suggested the localisation of inclusion bodies to the transitional epithelium of the cervix and Dunlop *et al.,* (1964), further proved a transmission link, through a study of urethral isolates taken from men with inclusion conjunctivitis, and men who were in contact with women who had inclusion conjunctivitis and genital tract infections.

Gordon and Quan, (1965) introduced the tissue culture technique for the isolation of *C.trachomatis*. The process involved the centrifugation of isolates from presenting patients upon a stationary phase monolayer of cells. Cycloheximide was later included in the cell culture medium to increase the isolation effectiveness. Cycloheximide inhibits deoxyribonucleic acid (DNA) and protein synthesis in eukaryotes, but not in prokaryotes such as chlamydiae, Ripa and Mardh, (1981); Reed *et al.,* (1981), to produce a stationary phase monolayer for the chlamydiae to infect. Gordon and Quan described the incubation of the cells that were stained to show inclusion bodies. Suitable antimicrobial agents were then used to control secondary infections.

Harper *et al.,* (1967) reviewed the egg culture and cell culture techniques and stated that 'yolk sac techniques are cumbersome, expensive and liable to spurious isolations due to cross infections'. In contrast the new cell culture technique after Gordon and Quan (1965) was observed to be more sensitive and the relative simplicity of the process made it suitable for examining large numbers of specimens. The Institute of Ophthalmology and London Hospital showed that cell culture could be used to obtain isolates of chlamydiae from the genital tract, eye, rectum and other sites.

During the 1970's and early 1980's chlamydiae research centred around testing of antimicrobial agents and antibiotics. The earlier work of Thygeson & Stone was arguably the basis of Gordon and Quan's collaborative study involving five laboratories, published in 1971. Using the standard cell culture

technique, described later, they reported isolation rates for NGU varied between twelve and forty-two per cent. Gordon and Quan attributed the differences to the collection techniques of specimens although this was only one factor. The data implied that the use of an endo-urethral curette to be more efficient than a metal cotton wool swab or a bacteriological loop. Philip, *et al.* (1971), confirmed Gordon and Quan's findings of the efficiency of the curette and reconfirmed correlations between ocular and genital infections between men with NGU and their female contacts.

Grayston and Wang, (1975) developed a microscopic immuno-fluorescence technique (micro-IF) allowing the separation of the *C.trachomatis biovar* into 15 different 'serovars' or serotypes. Prior to the application of micro-IF testing complement fixation (CF) tests had been used. But CF testing proved to be insufficiently sensitive to distinguish the *C.trachomatis* serovars from each other or to diagnose the oculo-genital infections that present with *C.trachomatis*. However, Complement fixation testing did prove useful in distinguishing the serodiagnosis of the LGV biovar and psitticosis.

Schleifer and Stackebrandt, 1983 established DNA-DNA re-association as a tool for the identification of chlamydial species and this lead to the identification of a new chlamydia species: *C.pneumoniae*, Grayston, (1989). In the same year the polymerase chain reaction was introduced as a technique and later helped to identify two further chlamydial species: *C.pecorum*, Fukushi and Hirai, (1993) and *C.suis*, Kaltenboeck *et al.,* (1993). The *C.suis* species lead to other chlamydial like isolates in swine being noted, Rogers and Andersen, (1996); Everett *et al.,* (1999). Although further testing of chlamydial variants in swine using sulfadiazines, found many to be resistant thus they cannot be classified as *C.trachomatis* under present criteria however, initially, *C.suis* was described as '*C.trachomatis*- like chlamydiae' in swine.

1.2 Chlamydia taxonomy and classification

The order Chlamydiales comprises four families although all species of Chlamydiae have been placed in the family Chlamydiaceae. It is at the genus level of chlamydial taxonomy that the more recent debate begins.

The intraepithelial inclusions of the trachoma causing agent, first described by Halberstaedler and von Prowazek in 1907, were identified as a protozoan and named *Chlamydia*, meaning cloak. Later, based on the organism's intracellular developmental cycle and the fact that Chlamydiae will not grow on conventional media, *Chlamydia* were classified as viruses.

After an extensive review of morphology, cytology, chemical nature and metabolism Chlamydiae were reclassified as bacteria due to their cellular organisation, mechanisms of macromolecule synthesis and cell division, and antibiotic susceptibility, Page, (1966); Oriel and Ridgway, (1982). This was confirmed by Schachter, (1990). Chlamydiae have been likened to Rickettsiales because they are also obligate intracellular parasites of eukaryotic cells. The differences between the two bacteria were highlighted and thus the first taxonomically valid use of the name *Chlamydia* according to Oriel & Ridgway (1982) was by Jones *et al.,* (1945).

In 1980 when the approved lists of bacterial names was published, Euzéby, 1980, the order Chlamydiales contained only one family: Chlamydiaceae, and one genus: *Chlamydia*. Within the genus *Chlamydia* were two species *C.trachomatis* and *C.psittaci* for members of the trachoma and psittacosis causing groups respectively, Page, (1968). This classification was based on sulphonamide sensitivity and glycogen accumulation in inclusions, both properties that distinguish *C.trachomatis* from *C.psittaci*. Each of these species

contained a variety of organisms but did not account for all strains of Chlamydiae, Dwyer *et al.,* (1972) and Darougar *et al* ., (1980).

Analysis of 16S rRNA genes demonstrated that *Chlamydia* are of eubacterial origin with no near relatives, Weisburg *et al.,* (1986). Taxonomic studies using genomic DNA confirmed that *Chlamydia* form a distinct taxonomic group (Frutos *et al.,* 1994). Although members of the genus *Chlamydia* display less than 10% inter-species DNA homology, (Cox 1998), their 16S rRNA gene sequences vary by less than 5% suggesting that all Chlamydiae have developed from a common ancestor, Thomas, (1995).

To date, Chlamydiae fall into the order Chlamydiales and have *'a distinct bacterial division that is deeply separated from other bacterial divisions'*, Schachter *et al.,* (2002). With advances in technology it is now possible to trace bacterial ancestry using DNA based classification techniques that were developed in the 1980's. DNA-DNA re-association, and sequence analysis are tools providing an insight into the molecular clock based on the 16S rRNA, which reflect the overall rate of evolutionary change for lines of microbial descent. DNA analysis techniques support the presence of eight chlamydial species in humans the most recent to be identified being: *C.pneumoniae*, Grayston, (1989) and *C.pecorum*, Fukushi and Hirai, (1993). Furthermore a ninth chlamydial species was identified in pigs, *C.suis* in 1993. It was described as '*C.trachomatis* -like *Chlamydia*' in swine, Kaltenboeck *et al.,* (1993). The nine species of *Chlamydia* currently identified are supported by phylogenic analysis, using the ribosomal operon and analysis of antigenicity, associated disease, phenotype, host range and biological data (Everett *et al* ., (1999); Pudjiatmoko *et al.,* (1997); Takahashi *et al.,* (1997); Kaltenboeck *et al.,* (1993)).

It has recently been proposed that the Chlamydiaceae family's single genus, *Chlamydia*, be divided into two distinct genera, the genus *Chlamydia* and the

genus *Chlamydophila,* (Everett *et al.,* 1999). The rationale for the split being based upon the following criteria: the species belonging to the genus *Chlamydia* should have at least 95% genetic sequence homology (Schachter *et al.,* 2002). Accordingly the included species of the *Chlamydia* genus are *C.trachomatis*, *C.suis* and *C.muridarum*. The species in the *Chlamydophila* genus are *C.abortus, C.caviae, C.felis, C.psittici, C.pecorum and C.pneumoniae*. However, this proposed change at the genus level has not been widely accepted by the scientific community, as evidenced by its use in very few scientific publications post 1999. The more favoured classification due to inconsistencies in the original proposal of Everett *et al.,* (1999) is that all nine species of Chlamydiae are placed under the single and traditional genus *Chlamydia*, (Schachter *et al.*, 2002).

1.3 Biological Differentiation within *C.trachomatis*

C.trachomatis is sub-divided in to two distinct biovars, the trachoma biovar and the LGV biovar. These biovars of *C.trachomatis* are distinguished in many different ways, but the most obvious distinction is the *in vivo* susceptibility of cells for infection by the two biovars. The strains of the trachoma biovar are also known as oculo-genital strains because of the areas they tend to infect *in vivo*. Oculo-genital strains infect columnar epithelial cells at mucosal surfaces. In contrast the LGV strains aggressively infect the lymphatic tissues ultimately causing systemic infection; this is due to a concentrated proliferation at the lymph nodes and subsequent distribution of infectious EB particles through the lymph vessels. This results in obvious differences in the clinical manifestation of the infections.

The LGV strains of *C.trachomatis* have an *in vivo* incubation period of between two and five days, although it has been reported difficult to assess this because of the transitory nature of the initial lesions (Oriel & Ridgway, 1982). These lesions can appear as small painless inguinal lesions, progressing, if untreated, to painful enlargement of the local lymph nodes and subsequent formation of buboes. The infection can also result in inflammation of the colon and rectum and excessive proliferation of intestinal and perirectal lymphatic tissue (Halse *et al.,* 2006;Spaargaren *et al.,* 2005). Untreated, the infection can lead to significant genital swelling, due to obstruction of the lymphatic vessels and can cause destruction of the rectum, forming rectal strictures and fistulae (Halse *et al.,* 2006;Mabey & Peeling, 2002). Ultimately, LGV can cause systemic infection and this is due to the initial infection proliferation at the lymph nodes and subsequent transportation of infection in the lymphatic fluids.

In contrast, the oculo-genital strains of the trachoma biovar are thought, *in vivo,* to only colonise columnar epithelial cells at mucosal surfaces this means that infections are usually localised to the ocular and/or genital areas and are

much less invasive when compared to the LGV strains. Ocular infections usually manifest as inclusion conjunctivitis, and are noticeable. Genital infections can, however, go unnoticed for years as in some cases there are little or no noticeable symptoms of illness or disease. It is possible to have urethritis and/or discharge, pain or lower abdominal discomfort but this is not necessarily the case in all infected individuals. However, untreated *C.trachomatis* infections are responsible for blindness, pneumonia and complications in pregnancy in some cases leading to abortion so it is important that any suspicion of infection is investigated thoroughly by a clinician.

Another notable difference between the trachoma and LGV biovars includes the increased virulence of the LGV strains when compared to the oculo-genital strains. The ability of the chlamydial strains of the LGV biovar to infect cells in culture without the need for centrifugation has been reported in the literature (Oriel & Ridgway, 1982). Both the LGV and oculo-genital chlamydial strains are cultivated in Buffalo Green Monkey Kidney cells (BGMK) at the University of Southampton. BGMK cells are infected when almost confluent in flasks and are treated with Cycloheximide to stop mitosis and put them in a stationary phase. The oculo-genital strains, however, unlike the LGV chlamydial strains require the inoculum to be centrifuged on to the cell monolayer. This is necessary to maximise the potential for infection of the monolayer. One possible explanation may be that elementary bodies close to the host cell sheet are normally repelled due to the net negative charges on the cell and elementary body surface (Oriel & Ridgway, 1982).

The observations in the literature suggest the increased virulence of the LGV chlamydial strains causes an increase in growth rate *in vitro* when compared to the chlamydial strains of the trachoma biovar. The LGV strains require two days growth before harvesting under optimal laboratory conditions whereas the trachoma biovar requires three days for optimum growth under the same laboratory conditions.

The two biovars of *C.trachomatis* are further divided into smaller groups based on variants in serological markers and are known as serovars. These serovars are identified with letters of the alphabet. The trachoma biovar has serovars A-K and the LGV biovar L1, L2 and L3.

The trachoma biovar has two variants in the disease it causes. There are the ophthalmic infections also known as trachoma, usually caused by serovars A-C, and there are the genital tract infections usually caused by serovars D-K. The *C.trachomatis* serovars were determined on the basis of serological immunoepitope analysis of the major outer membrane protein (MOMP) with monoclonal and polyclonal antibodies (Grayston & Wang, 1975; Molano *et al.,* 2004; Stephens *et al.,* 1982; Wang *et al.,* 1985).

The differences in the chlamydial serovars are mainly determined by the differences in the MOMP. These differences arise as a result of the four variable protein domains found within MOMP.

The MOMP of chlamydiae is one of the principal cell surface components (Stephens *et al.,* 1987) constituting over 60% of the total protein of the cell envelope (Caldwell *et al.*, 1981; Stephens *et al.,* 1986). MOMP is hypothesised to be a transmembrane protein with four exposed variable domains. These are labelled I to IV for ease of reference and are separated by a constant domain on either side (Stephens *et al.,* 1987). In addition, Stephens hypothesises that variable domains II and IV might account for antigenic diversity of chlamydiae on the basis that both domains are hydrophilic and contain considerable differences (Stephens *et al.,* 1987).

1.4 Epidemiology of *Chlamydia trachomatis*

C.trachomatis infection is of global significance and is the leading cause of tubal infertility and ectopic pregnancy in women and is the most common sexually transmissible bacterial pathogen world wide (Ngandjio *et al.,* 2003).

Cataract, glaucoma and trachoma are the three most common causes of blindness globally. In the less developed world *C.trachomatis* is associated mainly with ocular infection and trachoma, usually caused by serovars A-C. It is estimated that trachoma affects 146 million people, of whom nearly 6 million are blind. (Ejere *et al.,* 2004) *C.trachomatis* is easily treatable with antibiotics and the World Health Organisation (WHO) aims to eradicate trachoma blindness globally by 2020, although since this is a disease of poverty, this is dependent on economic growth as well as improvements in healthcare.

Trachoma in Europe was eradicated in the early 20th century; this was achieved through increased awareness of personal hygiene, better sanitation and reduced housing occupancy numbers through sociological advancement. (Rabiu *et al.,* 2005) Chlamydial ophthalmia neonatorum, chlamydial conjunctivitis and eye infections in adults, resulting from autoinoculation with genital *C.trachomatis* secretions is still seen in the developed world, however, these infections are identified early and treated before they progress and never manifest into trachoma.

In the developed world *C.trachomatis* causes genital tract infections, and is linked to complications such as pelvic inflammatory disease (PID) ectopic pregnancy and infertility. *C.trachomatis* genital tract infections are generally caused by serovars D-K.

Lymphogranuloma venereum (LGV) used to be an uncommon form of sexually transmitted disease usually observed in the tropics, it is caused by serovars L1, L2 or L3 of *C.trachomatis.*

There has been an increase in LGV infections globally, that coincides with an increasingly upward trend since the 1980's of world wide Chlamydial genital tract infections, and other sexually transmitted infections according to WHO published data. Alarmingly Chlamydial infections, although easily treatable with a course of antibiotics such as doxycycline and azithromycin, have been increasing significantly in numbers since 1996. This increase is irrespective of economic status or the availability of health care facilities. The WHO estimates 89 million new cases of *C.trachomatis* genital tract infection each year, world wide (Gerbase *et al.,* 1998), making it the most common sexually transmitted infection globally. However, this number is probably an underestimate given that the reporting of new cases and the use of laboratory diagnostic tests is inconsistent, particularly for men who are less likely to be screened than women (Schachter, 1999).

In the USA, there are 3 million new cases of *C.trachomatis* infection reported each year (Cates, Jr., 1999). In addition the USA saw an estimated rise of 20% in *C.trachomatis* infections between 2000 and 2002. *C.trachomatis* infection is recognised as the most common sexually transmitted infection reported to the US Centres for Disease Control and Prevention with the highest rates of infection being seen among adolescent females and young women. Approximately 5%-14% of screened females aged 16-20 and 3%-12% of women aged 20-24 present with a *C.trachomatis* infection in the United States. In addition to this, women are disproportionately affected by complications from genital tract chlamydial infections when compared to men, this is due to the often asymptomatic nature of the infection which may go undetected and persist for months or years (Golden *et al.,* 2000). These undetected chronic infections may eventually lead to pelvic inflammatory disease and possibly

result in infertility. This causes significant physical and psychological costs to the individual and financial implications for governments and individuals funding treatment. In 2004 the USA estimated direct medical costs of $249 million in the treatment of *C.trachomatis* infection presenting in adolescents and young adults. However, in 1998 an estimated $1.88 billion was spent by the USA for treatment of the resulting sequelae of *C.trachomatis* infection (Rein *et al.,* 2000).

An epidemiological study of 485 North American adolescents, who were HIV infected or high risk HIV uninfected individuals, showed that young age, a prior chlamydial infection, and multiple sexual partners significantly increases the risk for acquiring a *C.trachomatis* infection (Geisler *et al.,* 2004). During the follow up period of this study the group with highest incidence of presenting with more than one chlamydial infection were African American females less than 17 years of age who have had two or more sexual partners in the last 3 months and who have been diagnosed with a previous chlamydial infection. Of the 156 females with a diagnosis of chlamydial infection at least once during the study, 8% had a clinical diagnosis of pelvic inflammatory disease (Geisler *et al.,* 2004) with known sequelae of *C.trachomatis*.

The recognition that infection with chlamydia markedly enhances the risk for reproductive tract sequelae in women (including tubal infertility, chronic pelvic pain and ectopic pregnancy) sparked the development and expansion of wide spread chlamydia screening programs throughout the 1990's in many countries.

1.4.1 Review of the Epidemiology in Europe

There is a huge body of literature describing the epidemiology of *C.trachomatis*. This review will focus on the data from the Netherlands, Denmark, Sweden and the UK as comprehensive studies have been performed in these European countries and the lessons and conclusions are applicable to the UK situation.

1.4.2 The Netherlands

The Netherlands reported in 2003 a *C.trachomatis* infection prevalence of 23.57/100,000 population, this sharply rose in 2004 by 17.86% to 27.78/100,000 population, and again increased by 14.76% in 2005 to 31.88/100,000 population, according to the data published by the World Health Organisation [1].

The increasing prevalence led directly to a screening programme for the Netherlands to not only provide valuable data for this region but to correspond to data from other countries in Europe, such as Sweden, Denmark and the UK. A screening programme in the Netherlands would also help to provide an insight in to the incidence rates and distribution of *C.trachomatis* infections in Europe.

In 2003 as part of a study, a total of 21,000 15-29 year olds of both sexes were invited to partake in the first national population screening program for the Netherlands. The screening was organised by the Municipal Public Health Services (MHS) as a pilot scheme for a proposed routinely implemented screening strategy (Gotz *et al.,* 2006). This was a home based study. Of the 21,000 invited individuals 41% (8383) responded by sending in a urine specimen and a questionnaire. 11% (2227) returned a refusal card. The non respondents were from both lower risk and higher risk categories. The study

noted the overall chlamydial prevalence was 2% (Gotz *et al.,* 2006) this is a considerably lower percentage than those reported in the USA, UK (see section 1.4.5) and other European countries such as Sweden (Low & Egger, 2002).

It was shown by the results of this pilot scheme that the prevalence of *C.trachomatis* infection rose in concordance with the level of urbanisation of a test area. This was determined by comparing the rural areas that had a significantly lower *C.trachomatis* prevalence rate of 0.6% to the highly urbanised areas that had much higher 3.2% *C.trachomatis* prevalence rate (Van der Bij *et al.,* 2006).

The data showed that when the proportion of males to female in this case was taken into account, women had a higher prevalence than men, 2.5% and 1.5% respectively. The data showed that the prevalence of *C.trachomatis* infection was also associated with ethnicity (whether the participants were of Dutch origins or not), the number of sexual partners an individual had reported to have and the presence of symptoms, such as pelvic pain (Van der Bij *et al.,* 2006).

A further study by the MHS investigating *C.trachomatis* infections in Dutch 'high risk' populations, concentrated on multi- ethnic urban youth aged 15-29 years, via an outreach programme based in Rotterdam between September and December 2004. *C.trachomatis* testing was offered as an 'at home' based screening kit; the participation rate was 43% of the 556 individuals targeted.

The data showed that there was a *C.trachomatis* prevalence of 14.5% amongst sexually active individuals overall. There was a significantly higher prevalence of *C.trachomatis* infection recorded in females than males with results of 20.2% and 6.8% respectively. *C.trachomatis* positivity was reported

to be highest amongst individuals who attended training sessions at a vocational school (24.5%). The ethnic group with the highest *C.trachomatis* infection prevalence (17.5%) was the Surinamese/Antillean population who account for 27% of the non Dutch population of Rotterdam.

In total 48% of women and 13% of men reported having symptoms associated with an STI. 57% of women deemed sexually active had two to five partners and 10% had more than five lifetime partners. For the men these percentages were 27% and 66% respectively (Gotz *et al.,* 2006).

1.4.3 Sweden

The reporting of genital infection with *C.trachomatis* has been mandatory since April 1988 in Sweden, under the Communicable Disease Act (Jurstrand *et al.,* 2001). Screening programmes were initiated in Sweden during the mid 1980's to conduct opportunistic screening on unselected populations attending a variety of health care facilities. These populations were mostly composed of females (Herrmann & Egger, 1995).

It was noted 5-9 years after the screening programmes were initiated that the proportion of *C.trachomatis* positive samples tested by laboratories in these screening programmes had decreased by 2/3 (Herrmann & Egger, 1995).The reported incidence of genital chlamydial infection continued to decline in all of the Swedish counties until 1996. It has been suggested that this was due to contact tracing of index patients' current sexual partners and previous sexual partners, screening, and treatment of both asymptomatic men and women (Jurstrand *et al.,* 2001;Kamwendo *et al.,* 1998).

Recently, however increases in the prevalence of genital *C.trachomatis* infection have been noted in Sweden. This is in concordance with global increases in genital *C.trachomatis* prevalence. Clearly an increase in population testing, as has happened in Sweden due to increased social awareness, has influenced the reported genitally infected *C.trachomatis* positive prevalence. However, with this increase in population testing and social awareness accounted for, there has also been an increase in the proportion of *C.trachomatis* positive test results within the population from 4.1% in 1994 to 6.1% in 2000 (Low & Egger, 2002).

This resurgence of genital chlamydial infection in Sweden could be as a result of the failure to include males comprehensively in national screening

programmes. The opportunistic health programmes only reach health service users and it is noted that amongst sexually active individuals the majority of health service users are women (Low & Egger, 2002)

In Sweden only 20-25% of collected specimens for chlamydial testing come from men (Herrmann & Egger, 1995) therefore it is likely that there will be a circulating pool of undetected asymptomatic untreated chlamydial infection within the Swedish population (Low & Egger, 2002).

1.4.4 Denmark

In Denmark all positive cases of *C.trachomatis* are required to be reported to the Statens Serum Institute (SSI) in Copenhagen. A nationwide monitoring system, for the incidence and prevalence rates of *C.trachomatis* infection was initiated in January 1994 when it also became mandatory for physicians and clinics to report cases of *C.trachomatis* infection (Hoffmann 2001).

The data collected at the SSI in Copenhagen is hoped to firmly establish a nationwide monitoring scheme of the incidence of infection caused by *C.trachomatis* (Hoffmann 2001). The annual *C.trachomatis* positive laboratory confirmed infections in Denmark has stabilised to approximately 13,000 cases. (Hoffmann 2001). However, there has been a steady increase in the prevalence of *C.trachomatis* infection from 1989-2001.

Danish women aged 15-24 years and men aged 20-29 years are at the greatest risk of contracting a chlamydial infection. These findings are in concordance with the Netherlands, Sweden and the UK. The SSI collects analyses and compiles the *C.trachomatis* rates of prevalence within the Danish population and publishes the details quarterly. In addition, a more detailed report containing the incidence of *C.trachomatis* infection by age and county are published on an annual basis (Hoffmann 2001).

In 1999 women accounted for more than 70% of the *C.trachomatis* infected population in Denmark. Higher prevalence rates in women rather than men is a common observation of *C.trachomatis* epidemiological data and this disproportionate distribution reflects other similar findings within Sweden (Gotz *et al.,* 2006) the UK (Fenton *et al.,* 2001) and the Netherlands (Van der Bij *et al.,* 2006).

In a study of *C.trachomatis* in relation to age specific prevalence, the incidence rates of partners of index patients were also considered. It was reported that although the highest risk of *C.trachomatis* infection was in male partners of female *C.trachomatis* positive index patients, more than three times as many women at 5.8% were tested as sexual partners to *C.trachomatis* infected patients than men at 1.7%.

1.4.5 United Kingdom

There was a 43.7% increase in the diagnosis of chlamydia in the UK between 2001 and 2005 according to WHO published data [1]. Genital chlamydial infection is the commonest diagnosed sexually transmitted infection in the UK with 109,958 new cases reported in 2005, an increase of 5% from 2004 according to the Health Protection Agency, the reporting body for the UK. Implementation of preventative interventions, such as the national chlamydia screening programme in England, aimed at specific population risk groups have had little impact on the prevalence of *C.trachomatis* infection which keeps rising (Brown *et al.,* 2004). Similar strategies have been implemented in Wales, Scotland and Northern Ireland and all have drawn attention to the need for greater political will and investment in tackling STI prevalence in the UK (Brown *et al.,* 2004).

The greatest burden of disease is in young people under 25 years of age. Between 2001 and 2005 in individuals under 25 years of age there was a trend in women being diagnosed with chlamydial infection more so than men (3% in females) However, in individuals over the age of 25 the trend was reversed (3.1% in males). Being of a non-married status and reporting concurrency of two or more sexual partners in the past year were also independently associated with *C.trachomatis* infection. As the majority of women and a significant proportion of men diagnosed with this infection are asymptomatic, there is a significant pool of prevalent undiagnosed infection (Fenton *et al.,* 2001) which undoubtedly provides a reservoir for further *C.trachomatis* spread.

A study investigating the uptake of systematic postal screening for genital *C.trachomatis* and the prevalence of infection showed that 34.5% of the 19,773 males and females aged 16-39 years asked to participate collected and posted specimens. Postal reminders, telephone calls, home visits, and 'flagging' of patient records at local practices were employed to encourage participation.

The data showed that uptake in 'high risk' populations centered on 16-24 year olds, was 31.5% and was lower in men, young adults and practices in disadvantaged areas. The overall prevalence of chlamydial infection was lower in men than in women, 2.8% and 3.6% respectively and the prevalence of infection was found to be higher in people under the age of 25. In males under the age of 25 it was found to be 5.1% and in females under the age of 25 the prevalence was found to be 6.4%. It was concluded from this research that postal screening was feasible; however uptake was incomplete especially in deprived areas with a significant number of non-white residents. In addition women at a higher risk of infection were noted to be less likely to participate, this could mean that screening in this way might lead to wider inequalities in sexual health (Macleod *et al.,* 2005)

Opportunistic testing of women and their male partners has been suggested to reduce the burden of long term morbidity (Noone *et al.,* 2004). A Scottish study in three large family planning clinics during 2003 found there was no significant difference in *C.trachomatis* diagnosis between the clinics when opportunistic screening was implemented. The prevalence of *C.trachomatis* infection was 5.2% across the clinics. 98% of the women recruited accepted a test for *C.trachomatis*, those who did not partake in the study deemed themselves low risk because of a steady relationship or no current partner (Noone *et al.,* 2004). These results concurred with (Adams *et al.,* 2004)) who suggested that chlamydial prevalence estimates by age and setting could be used to inform screening strategies.

In another study from Scotland, 798 male army recruits were screened for *C.trachomatis* infection during their routine medical examination at Glencorse barracks. The prevalence of infection amongst British troop samples was 9.3% with similar infection rates across all age groups. Alarmingly 88% of *C.trachomatis* positive infections were asymptomatic, again highlighting the need for a national screening programme that targets males as well as females.

In contrast to studies based around genito-urinary medicine (GUM) clinics and other health settings, this study was based on a non self selecting population in terms of sexual activity and had 100% participation. The study also found that the rate of asymptomatic infections was higher than the 50% usually cited and shows the importance of including men as well as women in opportunistic screening for *C.trachomatis* infection (McKay *et al.,* 2003)

1.5 Project aims

Surveillance data are a basic prerequisite for monitoring, evaluating and improving sexual health policies. Basic epidemiological data on all STIs is planned to be routinely collected in almost all European countries. However with increases in *C.trachomatis* prevalence there is an urgent need for more comprehensive, detailed and appropriately time scaled data collection, analysis and reporting.

In the UK *C.trachomatis* infections rates are reported. The reported data however useful in documenting past annual UK *C.trachomatis* rates of infection and giving an indication of the size of the problem lacks the detailed molecular epidemiology that is required to determine a strategy to the significantly increasing rates of *C.trachomatis* infection currently being seen in the UK. This project is funded by the State of Jersey through the Wessex Medical Trust. There is significant concern about STI's within the island population. Therefore this investigation was planned as a pilot study to look at the molecular epidemiology of *C.trachomatis* in Jersey. This study was initially planned to focus on the varying island population of Jersey and the mainland city population of Southampton over the course of three years at a molecular epidemiological level to give an insight into the localised geographical distribution of different 'serovars' of *C.trachomatis*, and gain a better understanding of the spread of *C.trachomatis* infection.

Knowledge of the serovar distribution in the UK will allow comparison to that in other countries. As well as contributing to the molecular epidemiological insights into the UK population, this may lead to identification of risk factors associated with new and/or emerging strains. For example, if a significant new strain of a particular serovar is seen to emerge in a geographical area, the prevalence of that serovar in another geographical area and the 'traffic' of host

population between those areas presents the risk that the new strain will appear in the second geographical area. This is especially true if the strain possesses a selective advantage.

1.5.1 Primary objectives of the study

1. To develop methods for molecular typing of *C.trachomatis* by sequence analysis of nucleotide substitutions in the *ompA* gene using samples that have been processed in the BD Probe Tech protocol (Jersey samples)

2. To develop methods for the typing of *C. trachomatis* laboratory cultivated isolates, by sequence analysis of the *ompA* gene

3. To use these methods in a pilot study to evaluated the molecular epidemiology of chlamydial genital tract infection in Jersey

Both BD Probe Tech (used in Jersey) and Taqman diagnostic protocols (used in Southampton) detect the presence of the chlamydial cryptic plasmid. In October 2006 a mutated strain of *C.trachomatis* was described in Sweden (Ripa and Nilsson 2006). This strain was not detected by all of the conventional commercially available NAATs tests. Preliminary data suggests this new variant of *C.trachomatis* contained a deletion in the cryptic plasmid, and was first noticed because of an apparent 25% decrease in *C.trachomatis* infection prevalence between November 2005 and August 2006 in Halland county, southwest Sweden (Ripa & Nilsson, 2006)

In anticipation of the new variant *C.trachomatis* strain with a deletion in the cryptic plasmid reaching the UK and thereby affecting the positivity rate, two mutant isolates were sourced from Sweden for biological and molecular characterisation.

It is known that samples presenting with the 377bp deletion in the cryptic plasmid are detectable via the BD Probe Tech *C.trachomatis* detection system. However the Roche Cobas Taqman *C.trachomatis* detection system used at the Southampton HPA laboratory targets the region contained within the 377bp deletion. Therefore samples will be tested by both diagnostic methods to determine if there are any samples presenting with the 377bp deletion in the cryptic plasmid. In addition for confirmatory purposes it is intended that the Jersey and Southampton samples be screened for the new variant of *C.trachomatis* with a deletion in the cryptic plasmid that has been detected in Sweden using the same assay described by Ripa and Nilsson (2006). If there are samples with the 377bp deletion in the cryptic plasmid detected they will be *ompA* sequence genotyped to determine the isolate serovar identity and comparative analysis completed. The biology of the strain (isolate of known mutant from Sweden) will be studied to determine if there are any differences between it and a non mutated serovar E strain. This may then indicate selective advantages of the new mutated strain.

CHAPTER 2 MATERIALS AND METHODS

2.0. Culture of Chlamydiae

2.0.1 Bacterial strains

The following chlamydial strains were cultured for the isolation of genomic DNA,

Table 2.1 *Chlamydia trachomatis* strains sourced from the ATCC.

ATCC Number	Strain Name	Isolation Source
VR-902B	L2 434	Bubo from human with LGV
VR-348B	Trachoma type E strain BOUR	Adult with probable clinical inclusion conjunctivitis, 1959 Adult with clinical, active trachoma, California, 1959
VR-901B	L1 440	Lymph node from human with LGV
VR-903	L3 404	Lymph node from human with LGV

Table 2.2 *Chlamydia trachomatis* strains kindly provided by collaborators

Strain Name	Collaborator	Serovar
L2/434/I	Israel	L2
UCH1	University College Hospital, London, UK	L2b
SW2	Malmo, Sweden	E
SW3	Malmo, Sweden	E

2.1. Media

The following media was used for the culture of chlamydial host cell lines:
Dulbecco Modified Eagle Medium (DMEM) high glucose, 200mM glutamine (Invitrogen) supplemented with 10% foetal calf serum (Invitrogen)

The following media was used for the maintenance of chlamydial host cell lines immediately before infection with chlamydial strains:
Dulbecco Modified Eagle Medium (DMEM) high glucose, 200mM glutamine (Invitrogen) supplemented with 10% foetal calf serum (Invitrogen), 1g/mL of Cycloheximide (Sigma) 10g/mL gentamicin (Sigma).

Table 2.3 Contents of DMEM (Invitrogen)

Description	**mg/L**
INORGANIC SALTS	
$CaCl_2$ (anhydrous)	200.00
$Fe(NO_3)_3 \cdot 9H_2O$	0.10
KCl	400.00
$MgSO_4$ (anhydrous)	97.67
NaCl	6400.00
$NaH_2PO_4 \cdot H_2O$	125.00
VITAMINS	
D-Ca Pantothenate	4.00
Choline Chloride	4.00
Folic Acid	4.00
Myo-Inositol	7.00
Niacinamide	4.00
Pyridoxine HCl	4.00
Riboflavin	0.40
Thiamine HCl	4.00

Table 2.3 Contents of DMEM (Invitrogen) continued

AMINO ACIDS	
L-Arginine HCl	84.00
L-Cystine 2HCl	62.57
L-Glutamine	584.00
Glycine	30.00
L-Histidine $HCl \cdot H_2O$	42.00
L-Isoleucine	104.80
L-Leucine	104.80
L-Lysine HCl	146.20
L-Methionine	30.00
L-Phenylalanine	66.00
L-Serine	42.00
L-Threonine	95.20
L-Tryptophan	16.00
L-Tryosine $2Na \cdot 2H_2O$	103.79
L-Valine	93.60
OTHER	
D-Glucose	4500.00
Phenol Red (Sodium)	15.90
$NaHCO_3$	3700.00

2.2. Buffers, reagents and solutions

Buffers reagents and solutions were commercially sourced and made up with Ultra High Quality (UHQ) water, where required buffers, reagents and solutions were filtered for sterility using a 0.2µm syringe filter.

Table 2.4 Buffers, reagents and solutions

Description	Final concentration	Storage
Trypsin/EDTA (TE)	5% Trypsin / 20% ETDA	-20°C
Cycloheximide	1mg/mL	-20°C
Gentamicin	10mg/mL	-20°C
Dul A/Phosphate buffer solution (PBS)	1 tablet/100mL	Room temp
EDTA		Room temp
TAE	1x	Room temp
TBE	1x	Room temp
DMSO		2-5°C
4SP		-20°C
Orange G loading dye	6x	2-5°C

2.3 UHQ Water

Distilled water (dH_2O) produced by reverse osmosis using Elga water purification equipment was further purified by reverse osmosis to a resistance of 18 mega-ohms, yielded water of a higher quality (Ultra high quality, UHQ water), and this UHQ water was suitable for the preparation of solutions for nucleic acid manipulation. All general purpose and, solutions for nucleic acid manipulation were prepared with UHQ water.

2.4 Chlamydial host cell culture

Culture of Chlamydial isolates was completed in one of two host cell lines as follows:

Table 2.5 Chlamydial host cell lines

Description	Source
Buffalo Green Monkey Kidney Cell (BGMK)	Monkey
Mc Coy Cell	Mouse

2.4.1 Host cell storage

BGMK cell and McCoy cell stock cultures were stored in liquid nitrogen in 10% DMSO. Tri-annually (September, January and April) at the start of each academic term, cell lines would be thawed and resuscitated for research to continue.

2.4.2 Host cell resuscitation

1mL aliquots in 10% DMSO of either BGMK cells or McCoy cells were removed from the liquid nitrogen storage and thawed in the class II cabinet. The necks of tubes and bottles were wiped with 70% alcohol to prevent contamination. 9mL of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum was added to a universal tube and 1mL was used to re-suspend the thawed cells. The cells were then added to the 9mL of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum to make a total volume of 10mL. This was centrifuged at 1000rpm for 5 minutes to pellet the cells away from the DMSO, the supernatant was discarded. 1mL of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum was

added to the pellet and the cells were re-suspended this was added to a volume of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum (6-7mL T25 and 18-20mL T75) in a tissue culture flask. The caps of the tissue culture flasks were loosened before incubation at $37^{O}C$ in 5% CO_2.until confluence.

2.4.3 Chlamydia host cell passage

BGMK cells and Mc Coy cells were grown in T25 or T75 tissue culture flasks with non-filter caps (Greiner/NUNC), in DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum at $37^{O}C$ and 5% CO_2 for 24-48 hours or until confluent. All necks of flasks, and bottles were wiped with 70% alcohol to prevent contamination. Cells were removed from incubators, the media removed and washed three times with PBS to remove all traces of media and foetal calf serum which might inhibit cell trypsinisation from the flasks. A thawed 3mL aliquot of TE was added to the flasks and excess was poured off flasks which were replaced in the incubator for 5 minutes. Sterile flasks were labelled with date, cell line, researcher's initials and the split ratio. Flasks were filled to a maximum volume of 6-7mL (T25) and 18-20mL (T75) of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum including the calculated cell volume required for the passage. Cells were again removed form the incubator and inspected under the microscope at 20x magnification to ensure they had been 'freed' from the flask. Any cells still attached were 'freed' with a sharp tap of the flask into the palm of the hand. The appropriate amount of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum was added to the flask (dependent on the dilution factor required) however, this was not less than 4mL or in excess of 10mL to avoid too much of a froth. Remaining cells were released from the flask and clumps of cells were removed using a pipette with a bulb top fitted, by washing the media up and down the flask culture surface. A calculated cell volume was then added to each of the 'new' tissue culture flasks dependant on the split ratio. The caps of the tissue culture flasks were loosened before incubation at $37^{O}C$ in 5% CO_2.

2.5 Chlamydial culture

BGMK cells and Mc Coy cells were grown in T25 or T75 tissue culture flasks with non-filter caps (Greiner/NUNC), in DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum at 37^{O}C and 5% CO_2 for 24-48hours or until confluent. In a class II cabinet the necks of all flasks and bottles were wiped with 70% alcohol to prevent contamination. Cells were removed from incubators, placed in the class II cabinet, the media was removed and cells were washed three times with PBS to remove all traces of media and foetal calf serum as a precaution against contamination. DMEM high glucose, 200mM glutamine, supplemented with 10% foetal calf serum 1g/mL of Cycloheximide and 10g/mL gentamicin replaced the volume of media removed, minus the volume required for re-suspension of chlamydia for infection (dependent on the number of flasks to be infected and the percentage of infection required). All flasks were labelled with the strain of chlamydia being infected, the infection date, the initials of the researcher and the total volume in the flask once inoculation had been completed. The media volume required for the re-suspension of chlamydia was placed in a sterile universal tube and labelled 'inoculating media'. In the class I cabinet, chlamydiae for inoculation were thawed, and the required volume of inoculum added to the universal containing the inoculating media. The universal tube was tightly closed and vortexed for 30 seconds to ensure an even distribution. The inoculum was allowed to rest briefly to allow the settling of any aerosol and then the universal was opened with extreme care. The required volume of inoculum was added to each tissue culture flask (dependent on the size of the flask, and percentage infection required) using a pipette. Tips were discarded into hycolin solution after each inoculation.

2.5.1 LGV biovar strains

LGV strains do not require the added centrifugal assistance that the trachoma biovar isolates require. With this in mind flask caps were secured and then loosened on each inoculated tissue culture flask. The flasks were then placed carefully into plastic containers with lids. Each container was labelled with bio-hazardous tape and clearly labelled with 'Chlamydia' and the strain. The lids of the plastic containers had ventilation holes to allow the circulation of gases in the incubator. Plastic boxes containing tissue culture flasks were incubated for 48-72 hours or until the required infection rate was achieved (usually 70-90%).

2.5.2 Trachoma biovar strains

Infection of the trachoma strains requires centrifugation. Because this is completed external to the class I cabinet, special safety precautions were developed. Chlamydia inoculated tissue culture flasks were capped tightly and placed in re-sealable sandwich bags, (two flasks per bag (T25) one flask per bag (T75)). The sandwich bag was sealed and placed into the cradle of the centrifuge in the class I cabinet. The cradle was then carefully removed from the cabinet and placed within the centrifuge so that the necks of the tissue culture flasks were facing the centre of the centrifuge. Only flasks of the same size and a balanced volume/weight could be centrifuged at one time. The centrifuge lid was closed and the flasks centrifuged for 30 minutes at 1800rpm. Once completed the centrifuge cradles were replaced in the class I cabinet and the flasks were removed from the cradles and the sandwich bags. The caps on the tissue culture flasks were loosened and the flasks were then placed carefully into plastic containers with lids. Each container was labelled with bio-hazardous tape and clearly labelled with 'Chlamydia' and the strain. The lids of the plastic containers had ventilation holes to allow the circulation of gases in the incubator. Plastic boxes containing tissue culture flasks were incubated for 48-72 hours or until the required infection rate was achieved (usually 70-90%).

2.6 Chlamydia harvesting with glass beads

Chlamydia infected BGMK cell or McCoy cell monolayers were harvested at 48 hours post-inoculation or at 70% infectivity per strain. The tissue culture flasks were removed from the incubator and placed in the class I cabinet. The monolayers were detached from the flask by cell scraping and pooled. The pooled monolayer medium was then placed into a universal containing sterile glass beads. The medium and glass beads were vortexed for 30 seconds to break the cell membranes and the membranes of the inclusion bodies to release the EBs and RBs. The beaded mixture was removed from the glass beads and transferred to a second universal using a 1mL pastette. This was then centrifuged at 3,500rpm for 10 minutes to pellet both the cell debris and the EBs and RBs. The supernatant was discarded at this point. The pellet was re-suspended in 1mL of DMEM high glucose, 200mM glutamine supplemented with 10% foetal calf serum the re-suspended pellet was centrifuged at 1,000rpm for 5 minutes to pellet the cell debris. The supernatant was collected and measured and transferred to a bijoux, an equal volume of 4SP was added to preserve the inoculum during freezing. This harvested inoculum could be used to further infect cells or for DNA extraction. Aliquots were stored at -80^{o}C.

2.6.1 Chlamydia harvesting with glass beads for DNA

1x 25 cm^2 polystyrene culture flask of BGMK cell monolayer was harvested at 48 hours post-inoculation or at 70% infectivity per strain as outlined in the method above.

2.7 Chlamydia harvesting with a homogeniser

Chlamydia-infected BGMK cell or McCoy cell monolayers were harvested at 48 hours post-inoculation or at 70% infectivity per strain. In the class II cabinet 10 mL of 1:10 PBS was made and placed in a universal on ice. The tissue culture flasks were removed from the incubator and placed in the class I cabinet. The medium was removed from the cells and kept in a sterile 50mL tube. The cells were washed with 3mL of PBS, to remove all traces of media and foetal calf serum which could inhibit trypsinisation of the cells; this too was removed and kept in the 50mL tube. 3mL of TE was added to each flask and the flasks were incubated at 37^{O}C and 5% CO_2 for 5 minutes. Flasks were removed from the incubator and replaced in the class I cabinet; 3mL of saved media from the 50mL tube was added to each flask to inactivate the trypsin. The whole flask content was then collected into the 50mL tube. It was necessary to centrifuge 50mL tubes to significantly reduce the volume of potential inoculum. The 50mL tubes were tightly capped and placed inside special foam lined, screw top containers maximum of 4 tubes per container. These containers were then capped and fitted in the buckets of the centrifuge. Tubes were centrifuged at 3500rpm for 10 minutes, the supernatant was discarded into hycolin solution and the pellets were saved. Pellets were re-suspended into a paste using a vortex. A volume of 1:10 PBS was added (depending on the required final volume) to each of the re-suspended pellets and these were then pooled into a homogeniser mortar, the pestle was added and secured using a clamp. The potential inoculum was homogenised for 5 minutes and then removed from the homogeniser into a universal using a pipette. The cap was secured and the universal was centrifuged using the same safety precautions as before at 1000rpm for 5 minutes. The supernatant containing the chlamydiae was removed to another universal and the pellet containing the cell debris was discarded into hycolin solution. The supernatant was collected and measured; an equal volume of 4SP was added to preserve the inoculum during freezing. This harvested inoculum was aliquoted into cryo tubes and could be used to further infect cells or for DNA extraction. Aliquots were stored at -80^{o}C.

2.8 Separation of EBs and RBs with Urografin

Aliquots of inoculum containing 4SP to a volume of 8mL were removed from the storage at -80°C and thawed in a class I cabinet. EBs and RBs were pelleted by centrifugation at 35000rpm for 1 minute and the supernatant discarded into hycolin solution. The pelleted EBs and RBs were re-suspended in 3-4mL of PBS.

2.8.1 First gradient

In a 50mL tube make up a 20% urografin gradient to a volume of 35mL as follows and place in a Beckman ultra centrifuge tube.

2.6 20% Urografin gradient

Description	Quantity (mL)
PBS	28
Urografin	7

It is essential to have a balance tube of equal weight to the sample tube so it is necessary to double the volumes above to account for the balance tube. Layer the sample on top of the gradient. Place each ultra centrifuge tube in its holder and seal the lid of the sample tube. Because of the extremely infectious nature of the sample as much of the method as possible was completed inside the class I cabinet. The sample was ultra centrifuged at 25,000x*g* for 2 hours at 4°C.

2.8.2 Second gradient

In separate universals the 34%, 44% and 54% Urografin gradients were made as follows

2.7 34%, 44% and 54% Urografin gradients

Percentage gradient	Description	Quantity (mL)
34%	PBS	11.9
	Urografin	6.1
44%	PBS	6.2
	Urografin	4.8
54%	PBS	3.7
	Urografin	4.3

It is essential to have a balance tube of equal weight to the sample tube so it is necessary to double the volumes above to account for the balance tube. One Beckman ultra centrifuge tube was used and 18mL of the 34% gradient was added to the tube. 10mL of the 44% gradient was extracted in to a syringe with a large needle. The tip of the needle was placed at the bottom of the 34% gradient already in the Beckman centrifuge tube and the 44% gradient was added very slowly. There was an interface between the two gradients. The process was repeated with the 55% gradient and also with the other Beckman ultra centrifuge tube; both tubes were placed in the class I cabinet. The 20% gradient tubes on completion of centrifugation were removed from the centrifuge and the holders; the balance was discarded as was the supernatant of the sample tube. The pellet was re-suspended in 2mL of PBS and then layered onto the mixed gradient. The tube was replaced in the tube holder and sealed. As before, the tube was balanced using the balance tube and PBS, and then centrifuged at 25,000x*g* for 2 hours at 4°C. To remove the purified EBs it was necessary to transfer the ultra-centrifuged tubes to the class I cabinet and secure the sample tube in a clamp. The balance tube was again discarded. The

sample tube was viewed using the black band viewer card placed behind the clamped ultra centrifuge tube. Two bands were visible, the top band was the RBs and the lower band was the EBs. The EB band was removed from the gradient using a fine tip pastette at the gradient interface and transferred to a bijou. The EBs sample was further transferred to a screw cap centrifuge tube and the washing from the bijoux were also added. The tube was sealed and balanced as before and ultra centrifuged at 20,000x*g* for 30 minutes at 4°C. On completion of centrifugation the balance was discarded as was the sample supernatant and the pellet was re-suspended in 0.5mL of PBS and transferred to a labelled micro centrifuge tube. Samples were stored at -80°C.

2.9 Chlamydia positive clinical samples

Samples when received from Jersey in a BD Probe Tech sample tubes were re-aliquoted into 2mL screw top tubes with an O-ring, labelled and stored at -80°C until required. The details of each sample were anonymised and a sample batch number and individual sample number was assigned for identification during the research.

2.10 Genetic Analysis

2.10.1 DNA Extraction Methods

2.10.1.1 Extraction of chlamydial genomic DNA from laboratory cultivated strains

Genomic DNA was extracted from harvested inoculum frozen with 4SP using the Wizard Genomic DNA purification Kit (Promega). 1mL of inoculum was thawed and pelleted by centrifugation at 12,000rpm for 5 minutes. Chlamydial pellets were re-suspended in 480µL 50mM EDTA to this was added 120µL proteinase K at 1mg/mL (Fisher) to give a final concentration of proteinase K at 200µg/mL. This was incubated for 1 hour at 60^0C. Chlamydial pellets were re-suspended in 600µL of Nuclei Lysis Solution (supplied) by gently pipetting up and down. This was then incubated at 80^0C for 5 minutes to lyse the cells and then cooled to room temperature. 3µL of RNase solution (supplied) was added to the cell lysate and the tube inverted 2-5 times to ensure thorough mixing. This was incubated at 37^0C for 20 minutes and the sample then cooled to room temperature. 200µL of Protein precipitation solution (supplied) to the RNase treated cell lysate and vortexed vigorously for 20 seconds. The sample was incubated on ice for 5 minutes and the precipitated protein pelleted by centrifugation at 14,000rpm for 5 minutes. The DNA was precipitated in the supernatant solution by adding equal volumes of isopropanol and supernatant to a new 1.5mL tube this was gently inverted until the thread like strands of DNA became visible. The precipitated DNA was pelleted by centrifugation at 14,000rpm for 2 minutes. The supernatant was removed and discarded and the pellet was washed with 600µL of 70% ethanol, at room temperature. The ethanol was carefully removed using a needle and syringe and the tube was placed in a vacuum desiccator for 10 minutes to remove any residual ethanol. DNA pellets were re-hydrated with 200µL UHQ water.

2.10.1.2 Extraction of chlamydial genomic DNA from clinical samples via ion exchange gel purification

Genomic DNA was extracted from the Jersey clinical samples by pelleting 100µL of a selection of samples at 3,500rpm for 10 minutes the pellets were then washed with UHQ water for 30 minutes. To re-pellet the samples they were further centrifuged at 14,000rpm for 5 minutes the supernatant was removed and sample were re-suspended in 20µL of UHQ water. 2µL of proteinase K at 10mg/mL was added and the samples were incubated at 37^0C for 1 hour. Samples were centrifuged for 5 minutes 14,000rpm the supernatant was removed and the pellet was re-suspended in 200µL 5% Chelex in UHQ water. Samples were further incubated at 56^0C for 30 minutes to activate the Chelex resin. Samples were 'flick' mixed 10 times to ensure even distribution of the Chelex resin and then further incubated in boiling water for 8 minutes. The Samples were centrifuged at 11,610rpm to pellet any debris from the reaction the supernatant containing DNA was collected and stored in the fridge over night.

2.10.1.3 Extraction of chlamydial genomic DNA from clinical samples via MagNA Pure LC total nucleic acid isolation kit (Roche)

200µL of each sample was placed into the sample cartridge and all of the required reagents were loaded into the respective reservoirs.

Table 2.8 MagNA Pure reagent details

Label Colour	Reagent Name	Use
Black	Wash Buffer I	Removal of PCR inhibitors
Blue	Wash Buffer II	Removal of salt and proteins etc.
Red	Wash Buffer III	Removal of salts etc.
Green	Lysis Binding Buffer	Cell lysis, and the binding of the total nucleic acid
Pink	Proteinase K	Protein digestion
Caramel	Magnetic Glass Particles (MGP) suspension	Binding of the total nucleic acid
Yellow	Elution Buffer	Elution of pure total nucleic acid, dilution of eluates and reconstitution of proteinase K

Lysis binding buffer (supplied) is added to the sample, resulting in complete cell lysis and release of nucleic acids, nucleases are denatured at this point. Proteinase K (supplied) is added to the samples and the proteins are digested. Nucleic acids bind to the silica surface of the added MGPs due to the chaotropic salt conditions, isopropanol and the high ionic strength of the lysis binding buffer. MGPs with the bound nucleic acids are magnetically separated from the residual lysed sample. MGPs with the bound nucleic acids are washed repeatedly with wash buffer (supplied) to remove unbound substances like proteins (nucleases) cell membranes, PCR inhibitors such as heparin or haemoglobin, and to reduce the chaotropic salt concentration. Again MGPs with bound total nucleic acids are magnetically separated from the wash buffer containing residual sample debris. The purified nucleic acids are eluted at 70^0C from the MGPs in the wells of the elution cartridge, whereas the MGPs are retained in the reaction tip and discarded. (HPA Southampton VSOP36 and the manufacturer's instructions)

2.10.1.4 Extraction of chlamydial genomic DNA from clinical samples via QIAamp spin columns (Qiagen)

Into a 1.5mL tube 20µL of proteinase K (supplied) and 200µL of sample were placed. 1 sample volume of buffer AL (supplied) was added and the tube was flick mixed 10 times before being incubated at 56^0C for 10 minutes. Samples were briefly centrifuged to relocate drops in the lid and 200µL of 100% ethanol was added. The samples were pulse-vortexed for 15 seconds and centrifuged briefly to relocate drops in the lid. The sample was applied to the QIAamp spin column which had already been placed in to 2mL collection tube (supplied) and centrifuged at 8,000rpm for 1 minute to bind the DNA to the column. The column was placed in a clean 2mL collection tube (supplied) and the collection tube containing the filtrate as discarded. 500µL of buffer AW1 was added to the column and it was centrifuged at 8,000rpm for 1 minute. The column was placed in a clean 2mL collection tube (supplied) and the collection tube containing the filtrate was discarded. 500µL of buffer AW2 was added to the column and it was centrifuged at 14,000rpm for 3 minutes. The column was placed in a clean 2mL collection tube (supplied) and the collection tube containing the filtrate was discarded. The QIAamp spin column was placed in a 1.5mL tube and 200µL of UHQ water was added to the column this was incubated at room temperature for 1 minute and then centrifuged at 8,000rpm for 1 minute to elute the DNA.

2.11 Quantitation of DNA by agarose gel electrophoresis

The concentration of DNA in aqueous solution was determined by estimation against a known commercial ladder standard (Promega) after agarose gel electrophoresis (2.14.2)

2.12 Primer design and selection

The sequences were chosen according to the following criteria:

A) Primers should not contain any palindrome sequences,

B) The forward and reverse primers should not form any 3'-5' concatenates,

C) The primers should be 17-40 bases in length, and not more than 40% G-C rich

D) The termini of the primers should be complementary

E) The forward and reverse primers should have similar T_m values.

Primers were sourced from Sigma Genosys, supplied freeze dried and re-suspended upon receipt in UHQ water to a stock concentration of 100µM.

Table 2.9 Primer details for *ompA* specific amplification (Ngandjio *et al.*, 2003)

Primer Name	Orientation	Primer sequence
CT1	Forward	5' GCC GCT TTG AGT TCT GCT TCC TC 3'
CT5	Reverse	5' ATT TAC GTG AGC AGC TCT CTC AT 3'

Table 2.10 Primer details for *ompA* specific amplification (Jurstrand *et al.*, 2001)

Primer Name	Orientation	Primer sequence
P1	Forward	5' ATG AAA AAA CTC TTG AAA TCG G 3'
OMP2	Reverse	5' ACT GTA ACT GCG TAT TTG TCT G 3'

Table 2.11 Primer details for *ompA* variable domain specific amplification (Ngandjio *et al.*, 2003)

Primer Name	Orientation	Primer sequence
VD2	Forward	5' CAT ATG CAA GAT GCT GAG ATG TTT AC 3'
VD3	Reverse	5' AAA CTT GCT GAG ACC ATT TAA CTC C 3'
VD4	Forward	5' CTA TTG ATT ACC ATG AGT GGC AAG C 3'

2.13 Amplfication of chlamydial DNA by Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is an *in vitro* method of producing large amounts of specific DNA fragments from small amounts of DNA template. Oligonucleotide primers hybridise with the nucleotide sequences on complementary strands at each end of the DNA fragment to be amplified. Cyclical polymerisation by thermostable DNA polymerase produces millions of identical copies of the DNA fragment.

Amplification of chlamydial template DNA for visualisation was completed by PCR. All sample batches received were initially screened for *C.trachomatis* positivity using *ompA* specific primers shown in tables 2.9 and 2.10 (Jurstrand *et al.*, 2001; Ngandjio *et al.*, 2003). This initial screening amplified the whole of the *ompA* gene which was visualised and characterised via agarose gel electrophoresis. Because often more amplified material was required than that resulting from the initial screening those samples confirmed to be *C.trachomatis* positive were amplified to provide enough material for visualisation, characterisation and recovery via gel extraction. Amplified material needed to be of a quantity that after gel extraction the sample could be again visualised and characterised via agarose gel electrophoresis to ensure the required quantity be added to reactions for automated DNA sequencing.

2.13.1 PCR reaction components

Bioline Bio-X-Act commercial kits were used for amplification of PCR products. A total reaction volume of 20µL was used per sample. The Bio-X-Act PCR kit (Bioline) and primers (Sigma) were stored at -20^{0}C between uses. The reaction was mixed in a 0.2mL double-snap-cap microcentrifuge tube
The PCR components per reaction were

2µL 10x Taq Buffer

1.4µL $MgCl_2$, (50mM)

0.4µL dNTPs 10mM (Amersham Pharmacia Biotech Inc.)

2µL forward primer (10mM)

2µL reverse primers (10mM)

0.25µL (1 unit) Bio-X-Act thermostable DNA polymerase (Bioline)

1µL Template DNA

10.95µL UHQ water

2.13.2 Thermocycling conditions

All PCR reactions were performed in a Gen Amp PCR system 9700 (Applied Biosystems), and thermal cycling criteria were selected on the following basis:

A) Tubes were placed in a thermocycler preheated to 95^0C to reduce false priming.

B) Denaturation
95^0C for 1 minute in all reactions to reduce false priming.

C) Annealing of the primers
The annealing temperatures were selected on the basis of the T_m of the primer pair, 10^0C below the lowest T_m of the primer pair. In a series of experiments this was gradually increased towards the T_m to maximise the annealing to the template DNA and therefore the amplification. The reaction mix was held at the optimised annealing temperature for 1 minute

D) Extension of the DNA strands
The temperature at which the Taq polymerase was allowed to polymerise the extension of the new DNA fragment was always 72^0C. The time allowed for this extension was 1 minute since the desired PCR products were approximately 1.1 kb in length.

E) Final extension
After 35 cycles of PCR, a final extension step was completed at 72^0C for 10 minutes to complete the extension of any uncompleted PCR products

The temperature of the thermocycler was altered by a rapid ramp. The reactions were held at 10^0C after thermo-cycling was complete until removal from the thermocycler.

2.14 PCR amplicon detection, characterisation and quantification via agarose gel electrophoresis

Agarose gel electrophoresis was used for the separation, visualisation characterisation and quantitation of DNA fragments in comparison with a commercial standard (Promega) as a method to enable the purification of specific DNA fragments. The concentration of agarose used to prepare gels for the horizontal submerged electrophoresis was determined by the size of the DNA fragments to be resolved. The majority of samples were successfully resolved using 1% agarose gels, however certain samples could not be resolved at this concentration and then the concentration was taken from the information outlined in the table below

Table 2.12 Guide-lines for agarose gel concentrations

Agarose	Range of DNA resolved (kb)
0.6%	20-1
0.7%	10-0.8
0.9%	7-0.5
1.2%	6-0.4
1.5%	4-0.2
2.0%	3-0.1

50x, pH 8.3, Tris-Acetate-EDTA (TAE) stock (Promega) was diluted to a working concentration of 1x with UHQ water. The 1x TAE working solution was used as the electrophoresis buffer for horizontal submerged agarose gels and, as a solvent to dissolve the agarose powder for the gel mix.

The agarose gels were cast in perspex trays (Bio-Rad) of various sizes. The ends of the trays were sealed with masking tape and placed on a level surface. A comb was used to form sample wells and was placed in the tray approximately 2cm from the top of the gel and 2mm above the base of the tray.

Agarose (Fisher) was dissolved in 1x TAE, by boiling in a microwave oven. The sizes of gels used and the volume of agarose of melted agarose required are outlined below.

Table 2.13 Gel dimension specifications

Gel dimensions	Volume of melted agarose required
6.5cm x 10cm	30mL
15cm x 10cm	100mL
15cm x 20cm	100mL

Agarose was cooled to approximately 50°C, and poured into the gel mould without allowing bubbles to form. The gels were allowed to set and rest for a minimum of 1 hour before use to allow syneresis to occur. Once set the masking tape and comb were removed and the gel was submerged fully in 1x TAE buffer ready for loading and electrophoresis.

Samples to be loaded contained 1x Orange G loading buffer and were loaded under the 1x TAE electrophoresis buffer. Gels were routinely electrophoresed at 100 volts for 1 hour.

2.14.1 Visualisation of DNA

Once the Orange G dye front within the sample loading buffer had migrated to 1cm from the end of the gel, the agarose gels were removed from the electrophoresis tank and casting tray and were placed in 0.5µg/mL ethidium bromide (Fisher) to a final volume of 200mL. Gels were stained for 1 hour. Excess ethidium bromide was washed away with UHQ water. The DNA was visualised by ultra-violet illumination using a transiluminator (Ultra-violet products Inc) at 302nm. Photographic gel images were taken using Polaroid 667 film.

2.14.2 Estimation of DNA size and concentration

The sizes of DNA fragments were estimated by comparing their relative mobility through the gel matrix against DNA fragments of a known size. The DNA samples were routinely comparatively analysed using a 1kb (Promega) and 1kb plus DNA ladder (Invitrogen) as a standard for both quantification of DNA and characterisation of the separated band sizes.

2.15 1kb ladder

The DNA 1µg/µL was supplied in a buffer containing 10mM Tris-HCl (pH 7.4), 1mM EDTA

Table 2.14 DNA ladder band sizes 1kb

Size (bp)	Size (bp) cont.	Size (bp) cont.
10,000	3,000	1,000
8,000	2,500	750
6,000	2,000	500
5,000	1,500	200
4,000		

2.16 1kb plus ladder

The DNA 1µg/µL was supplied in a buffer containing 10mM Tris-HCl, pH 7.5; 10mM EDTA, (pH 8.0); 0.06% XCFF; 0.6% tartrazine; 5% glycerol; 5mM NaCl

Table 2.15 DNA ladder band sizes 1kb plus

Size (bp)	Size (bp) cont.	Size (bp) cont.	Size (bp) cont.
12,000	7,000	2,000	500
11,000	6,000	1,650	400
10,000	5,000	1,000	300
9,000	4,000	850	200
8,000	3,000	650	100

2.17 Purification of amplified products

Before sequencing, the amplified DNA fragments were purified from the agarose gel to remove contaminating nucleotides and primers. Purification was accomplished using the Gene-elute gel extraction kit (Sigma) or the Qiaquick gel extraction kit (Qiagen).

2.17.1 Gene-Elute gel extraction kit (Sigma)

Stained and photographed gels were transferred to a short wave UV light box and DNA fragments of interest were excised with a clean sharp scalpel. Excised gel fragments were weighed and recorded. Gel extractions were completed using commercially available Gene-Elute binding columns (Sigma). Weighed excised gel fragments were dissolved in three times gel volume of solubilisation solution provided with the kit, in a 60°C water bath. 1x gel volume of isopropanol was added to the solubilised gel solution and vortexed for 10 seconds. Binding columns were prepared with 500µL of column preparation solution provided with the kit and the collection was discarded. The solubilised gel solution and isopropanol mixture was loaded into the binding column 700µL at a time and centrifuged at 14,000rpm for 1 minute. All collections were discarded. Columns were then washed with 700µL of the wash solution provided and centrifuged at 14,000rpm for 1 minute. The collection was discarded and columns were then centrifuged for a further 1 minute at 14,000rpm. Columns were transferred to a new collecting tube and placed in a vacuum for 30 minutes. DNA was eluted using 50µL of the elution solution provided with the kit. This was warmed in a water bath at 65°C for 10 minutes prior to use. A 5µL sample of eluted DNA was electrophoresed as stated above to visualise separation and quantitate the recovered DNA fragments. The DNA samples were comparatively analysed using a 1kb or 1kb plus DNA ladder

2.17.2 Qiaquick gel extraction kit (Qiagen)

Stained and photographed gels were transferred to a short wave UV light box and DNA fragments of interest were excised with a clean sharp scalpel. Excised gel fragments were weighed and recorded. The excised DNA band to be purified was mixed with three volumes of buffer QG (supplied) and incubated at 50°C for 10 minutes until the gel slice had completely dissolved. One gel volume of isopropanol (Fisher) was then added to the sample and mixed; the

sample was then loaded into the Qiaquick spin column. The product was bound to the column by centrifugation at 14,000rpm for 1 minute. Excess agarose was removed by the addition of 0.5mL of buffer QG (supplied) and centrifugation at 14,000rpm for 1 minute. The column was washed by the addition of 750µL buffer PE (supplied) and left for 5 minutes before centrifugation at 14,000rpm for 1 minute. The column was dried by a further 1 minute centrifugation at 14,000rpm and vacuum dessication for 10 minutes to remove all traces of residual ethanol from the buffer. The purified product was eluted in 50µL UHQ water by standing the column for 1 minute after the addition of the UHQ water. The tube was then centrifuged at 14,000rpm for 1 minute. A 5µL sample of eluted DNA was electrophoresed as stated in section 2.14 and then stained to visualise separation and quantify the recovered DNA fragments. The DNA samples were routinely compared against a 1kb or 1kb plus DNA ladder (Promega)

2.18 Automated DNA sequencing

Automated sequencing was performed on an Applied Biosystems model 377 DNA Sequencing System with an AB1 Prism Big Dye v1.1 Terminator Cycle sequencing Ready Reaction kit. The performance of the kit relies on four fluorescent dye-labelled di-deoxynucleoside terminators (ddNTP) and employs a thermostable DNA polymerase (Ampli Taq DNA Polymerase, ABI Prism) to determine DNA sequences by the chain termination method (Sanger *et al.*, 1977). The Sanger method of DNA sequencing can be summarised in three steps:

1) Annealing of the primer DNA to the template DNA,

2) Labelling of the DNA during polymerisation

3) Termination of the chain extension.

Each sequence determination is carried out as a single reaction, containing all four deoxynucleoside triphosphates (dNTP) in ratio to the fluorescently labelled ddNTP. The ddNTP lacks the necessary 3'-OH group required for chain elongation, thus the oligonucleotide chain is terminated selectively at the A, C, G, or T depending on the particular ddNTP incorporated. The resulting fragments have a common origin but are terminated at different nucleotides along the extension product and therefore can be separated according to size by high - resolution denaturing gel electrophoresis. In automated sequencing, a fluorescent dye label specific to A, C, G or T is incorporated into the DNA along with the terminating base. A laser scans across a region of the gel during the electrophoresis process, detecting the fluorescently labelled DNA passing through this region. Because each base is labelled with a different dye all four termination reactions can be performed in a single tube reaction and electrophoresed in the same lane of the gel.

2.18.1 Materials

To avoid any fluorescent contamination, reagents of the highest quality and UHQ water were used. All glassware was cleaned carefully by treatment with concentrated nitric acid and then rinse thoroughly with UHQ water. A ready to use stock solution of 4.8% Gene Page Plus was purchased from Amresco, containing 6M urea and 1x TBE buffer, and was stored at 4^0C. A 5x liquid concentrate of TBE buffer at pH 8.3 was purchased (Amresco) and stored at room temperature. A 1x working dilution was used as the electrophoresis buffer for the vertical acrylamide sequencing gels.

2.18.2 Sequencing reactions

The sequencing reactions were performed using an Applied Biosystems Gen Amp PCR system 9700. The DNA sequencing kit Applied Biosystems Big Dye v1.1 terminator cycle sequencing kit was stored at -70^0C until required, then at -20^0C between uses. High quality template DNA is required for sequencing. The table below shows the amount of template DNA used in a cycle sequencing reaction.

Table 2.16 Amount of template DNA required per sequencing reaction

Template PCR product	Quantity
100-200bp	1-3ng
200-500bp	3-10ng
500-1000bp	5-20ng
1000-2000bp	10-40ng
>2000bp	40-1000ng

In general the higher quantities of DNA template used gave higher signal intensities. The reaction was mixed in a 0.2mL double-snap-cap microcentrifuge tube with 6µL of the supplied premix (Big Dye Terminator v.1.1 ready reaction mix, ABI Prism) 2µL of supplied dilution buffer (ABI Prism) and 3.2pmol of primer in a total volume of 20µL. The primers were generally 18-30bp oligonucleotides. The tubes were placed in a thermal cycler preheated to 96^0C to reduce false priming. Denaturation was at 96^0C for 10 seconds, annealing T_m -10^0C for 5 seconds and extension at 60^0C for 4 minutes, for 25 cycles. The temperature of the thermocycler was altered by a rapid ramp. The reactions were held at 10^0C after thermo-cycling was complete. The extension products were purified by, ethanol precipitation, rinsing with 70% ethanol and vacuum desiccation for 15 minutes. The DNA pellet could be stored at -20^0C for up to 4 weeks.

2.18.3 Gel preparation

The Gene Page Plus gel was cast between two glass plates separated by spacers of 0.2mm constant thickness. To avoid fluorescent contamination the glass plates were washed thoroughily with detergent Alconox (Merck Ltd.) rinsed thoroughily with hot water, cold water, UHQ water and then industrial methylated spirits and air-dried. The plates and spacers were assembled and secured in the supplied gel cassette, then placed on a level surface to form a mould for the gel. 150µL of 10% APS solution and 15µL of TEMED solution were added to 25mL of Gene Page Plus solution to initiate polymerisation. A single flat comb (PE biosystems) was inserted into the top of the gel. The gel was then allowed to polymerise for 2 hours.

2.18.4 Gel electrophoresis

Once the gel had set, the comb was carefully withdrawn and the laser-scanning region cleaned by wiping with industrial methylated spirits and lint free Kim wipes. The plates were scanned to ensure the absence of any background fluorescence. The lower buffer reservoir on which the bottom edge of the plates rests was placed in the electrophoresis chamber. The plates were secured and the upper buffer chamber placed at the top of the plates. A 36 well disposable sharks tooth comb (PE biosystems) was then inserted carefully between the plates so that the teeth were inserted to a depth of 0.2mm and the buffer reservoirs filled with 1x TBE. The sample pellets were re-suspended in 4µL of loading buffer (5 deionsed formamide: 1 blue dextran). The samples were denatured by heating at 95^0C for 3 minutes and 2µL loaded into alternate wells that were pre-washed with 1x TBE. Electrophoresis was carried out for 12 hours at 2,500 volts, 40 mAmps, 30 watts and at 40^0C. The sequence data were automatically collected and analysed by computer (Applied Prism and DNA star software).

2.19 Computer analysis

DNA sequence editing, translations, open reading frame locations and sequence alignments were performed using Lasergene 6, Artemis and ACT software. Each of these computer packages allows the user to visualise sequence in a FASTA format complete the necessary editing of the FASTA sequence and visualise protein translations, annotate the file with open reading frame locations using the mapped start and stop codons and sequence alignments with either other FASTA sequences produced in-house or available via the internet such as those located on the NCBI website.

CHAPTER 3 RESULTS JERSEY

3.0 Samples available for this study from Jersey

This was a new project where liaison between the Jersey and Southampton diagnostic laboratories was necessary to coordinate sample collection. Table 3.1 shows the number of *C.trachomatis* positive samples detected in Jersey via the BD Probe Tech assay over a 12 month period (January-December) for 2005 in Jersey.

Table 3.1 2005 annual prevalence of *C.trachomatis* on Jersey

Total number of samples screened for *C.trachomatis*	Total number of *C.trachomatis* positive samples
3677	211

The aim of the project was centred on the epidemiological distribution and prevalence of *C.trachomatis* within a UK population. Therefore it was necessary to establish a systematic way of collecting the BD Probe Tech *C.trachomatis* positive samples from Jersey. Batches of approximately 50 BD Probe Tech *C.trachomatis* positive samples were collected at the diagnostic laboratory, in Jersey and anonymised. The samples were then couriered to the Southampton diagnostic laboratory. 50 BD Probe Tech *C.trachomatis* positive samples per batch was considered a feasible number of samples to store after diagnosis and individually packaged before transportation. On arrival in Southampton the date was recorded and this was used to identify the batch. All individual samples were given a sample number so that patient anonymity could be protected. Table 3.2 shows the number of samples collected over the first year of this project.

Table 3.2 BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2005

Batch identification number	Total number of samples	Total number of male samples	Total number of female samples
01.01.05	23 clinical samples	6	17
12.10.05	49 clinical samples	8	41
Total	72 clinical samples	14	58

Table 3.3 BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2006

Batch identification number	Total number of samples	Total number of male samples	Total number of female samples
14.02.06	48 clinical samples and 2 controls	6	42
31.05.06	60 clinical samples and 3 controls	25	35
27.10.06	57 clinical samples and 3 controls	25	32
Total	165 clinical samples and 8 controls	56	109

Table 3.4 BD Probe Tech determined *C.trachomatis* positive samples received from Jersey during 2007

Batch identification number	Total number of samples	Total number of male samples	Total number of female samples
20.04.07	49 clinical samples	22	27
04.07.07	47 clinical samples	13	34
Total	96 clinical samples	35	61

In Jersey, collected *C.trachomatis* positive samples remained in the BD Probe Tech specific sample collection and processing tubes, this minimised the diagnostic laboratory staff duties. Approximately 2mL of sample remained after processing via the BD Probe Tech diagnostic assay. This meant that there was a finite volume of sample for the planned subsequent research applications. Therefore, it was decided that should any clinical sample collection swabs be found in the *C.trachomatis* positive samples that they should remain to give the highest possible chance of *C.trachomatis* DNA template being successfully amplified in the planned subsequent investigations based in the Southampton research laboratories.

3.1 Development of *ompA* PCR using template DNA from BD Probe Tech method

The clinical samples from Jersey received at the Southampton laboratory had been subjected to diagnostic testing via the BD Probe Tech method. This is a qualitative method of *C.trachomatis* detection utilising strand displacement amplification (SDA) and fluorescence resonance energy transfer (FRET). The assay is based on the simultaneous amplification and detection of *C.trachomatis* plasmid target DNA using specific amplification primers and a fluorescently labelled detection probe. The samples collected for *C.trachomatis* diagnostics are typically endocervical swabs, male urethral swabs, and urine specimens;

these were from individuals who were symptomatic or asymptomatic for *C.trachomatis* infection due to the nature of the screening program in Jersey.

In the BD Probe Tech, SDA technique, *C.trachomatis* detection assay components include reagents that preserve nucleic acids. Whilst the actual components of the assay remain commercially confidential, it was thought that these BD Probe Tech samples may also have contained suitable DNA template ready for PCR reactions. The BD Probe Tech *C.trachomatis* detection system is designed for the direct qualitative detection of *C.trachomatis* DNA in endocervical swabs, male urethral swabs and urine specimens. This system is not designed or optimised for subsequent PCR investigations, therefore some optimisation was considered necessary.

The samples deemed *C.trachomatis* positive by BD Probe Tech were given a 'Method Other Than Acceleration (MOTA) score' of equivalent or above 2000. Negative samples were those with a score below the 2000 cut off. The MOTA score is the metric used by BD Probe Tech to determine the magnitude of the fluorescent signal generated by the SDA reaction. However the MOTA score is not indicative of the quantity of *C.trachomatis* organism in the clinical samples. This is because the BD Probe Tech target is plasmid DNA not genomic DNA and *C.trachomatis* are known to have several plasmids per genome copy (Pickett *et.al* 2005) the quantity of organism in the samples could be determined if the target was genomic DNA however, these detected organisms may not all be viable for culture.

Based on a preliminary review of methods by Dr Peter Marsh (personal communication), the method described by Ngandjio *et al.,*(2003) was chosen and an *ompA* specific PCR was developed. The method described by Ngandjio *et al.,* (2003) was chosen based on the following criteria:

A) The primers CT1 and CT5 amplified a PCR product that encompassed the entire *ompA* gene of *C.trachomatis*.

B) In addition, the method was designed to estimate the prevalence of *C.trachomatis* infection within a population, and to 'genotype' individual samples.

This information had then been used by Ngandjio *et al.,* (2003) as we also hoped to use our findings, to analyse the relationship between clinical signs and genotypes, to identify risk behaviour and to demographically analyse sexual networks. Ngandjio *et al.,* (2003) had been cross referenced with other work in the field (Hsu *et al.,* 2006) and was thought to be a credible starting point for development.

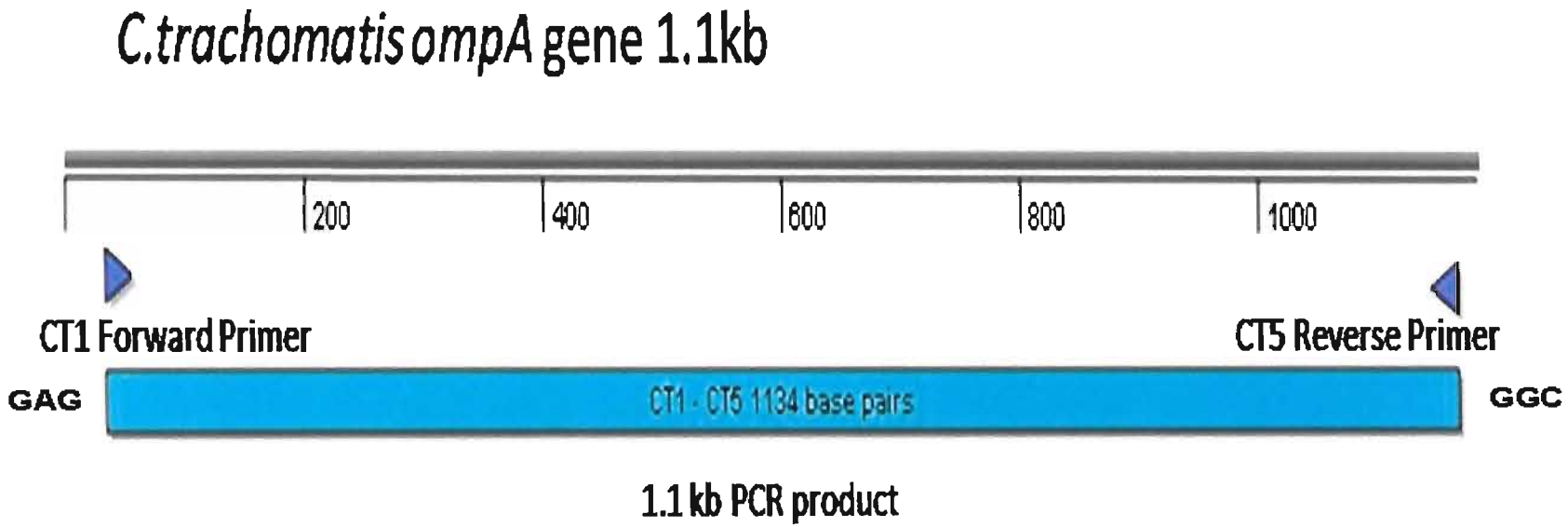
Ngandjio *et al.,* (2003) described initial *C.trachomatis* positive detection via the Roche Cobas Amplicor plasmid PCR test. This test is a Taqman based NAAT that simultaneously amplifies the target plasmid DNA with *C.trachomatis* specific primers and detects the cleaved dual fluorescent dye-labelled oligonucleotide detection probes allowing the detection of the 'reported' *C.trachomatis* specific amplified product.

Ngandjio *et al.,* (2003) described the isolation in cell culture of Roche Cobas Amplicor determined positive *C.trachomatis* samples. Bactopick swabs were taken from students in Cameroon and submersed in 1.5mL of 2-sucrose phosphate (2SP) transport medium. The medium was aliquoted into 3 microtubes and frozen at -80^{0}C within 3 hours of collection. 3 months after collection samples were tested for *C.trachomatis* using the Roche Cobas Amplicor technique. Positive samples were cultured in M^{c}Coy cells treated with Cycloheximide and the DNA harvested. The *ompA* gene was then amplified with

the CT1 and CT5 primers. Table 2.9 in the method section shows the details of the CT1 and CT5 primers.

A map indicating the location of the primers used in this method is shown in Figure 3.1

Figure 3.1 **CT1 and CT5 primer location and PCR product size (Ngandjio *et al.*, 2003)**



The aim in the initial stage of this project was the development of a successful *ompA* specific PCR method to produce PCR products from the BD Probe Tech *C.trachomatis* positive samples from Jersey. Any *ompA* specific PCR products would later be sequenced for the purpose of genotyping.

The study was initiated by repeating the PCR conditions outlined in the method described by Ngandjio *et al.,* (2003). However, all PCR development was completed utilising DNA extracted from laboratory cultivated LGV strains before the methods were tested on the BD Probe Tech *C.trachomatis* positive samples collected from Jersey. This was because these samples were unique, limited in volume and irreplaceable. In addition, the use of a constant control DNA allowed changes and improvements in the methodology to be maintained.

3.1.1 Laboratory cultivated LGV strains

The genomic DNA was extracted from the L2 isolates using the Wizard Genomic DNA purification Kit (Promega) as in section 2.10.1.1 of the methods. DNA from two strains L2/Amsterdam/Proctitis and L2/434/Bu/I sent to us by our collaborators in Israel was used for the initial work as detailed as follows.

3.1.2 L2/Amsterdam/Proctitis DNA concentration estimation

5µL of the L2/Amsterdam/Proctitis DNA was electrophoresed on an agarose gel as described in sections 2.14-2.17 of the methods, for visualisation, characterisation and quantification.

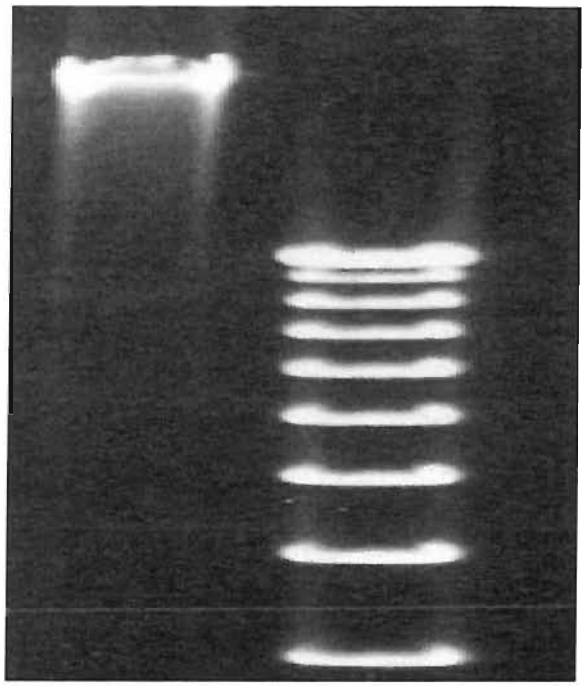


Figure 3.2: L2/Amsterdam/Proctitis original genomic DNA

Lane 1: L2/Amsterdam/Proctitis original genomic DNA Lane 2: 1kb plus ladder

The DNA concentration was estimated to be 80ng/µL by comparison to 1kb plus ladder (Figure 3.2)

A dilution of the original DNA sample was made with UHQ water to give a final DNA concentration of 1.6ng/µL. This was used as a test 'positive' control for the BD Probe Tech *C.trachomatis* positive samples from Jersey and for use as a DNA template for future optimisation reactions. This shall be referred to from here forth as L2/Amsterdam/Proctitis (1.6ng/µL).

3.1.3. L2/434/Bu/I DNA concentration estimation

5µL of the L2/434/Bu/I DNA was electrophoresed on an agarose gel as described in sections 2.14-2.17 of the methods, for visualisation, characterisation and quantification.

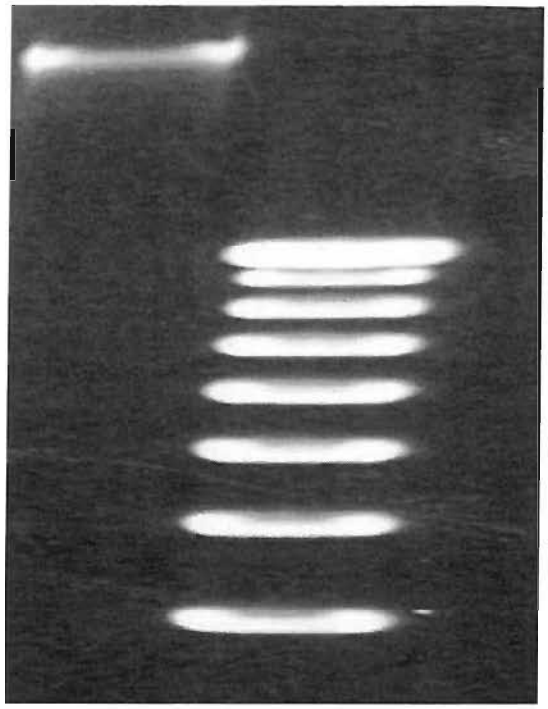


Figure 3.3: L2/434/Bu/I original genomic DNA

Lane 1: L2/434/Bu/I original genomic DNA Lane 2: 1kb plus ladder

The DNA concentration was estimated to be 60ng/µL by comparison to 1 kb plus ladder (Figure 3.3)

Dilution of the original DNA sample was made with UHQ water to give a final DNA concentration of 1.6ng/µL. This was used as a DNA template for future optimisation reactions and shall be referred to from here forth as L2/434/Bu/I (1.6ng/µL).

3.1.4 CT1 and CT5 primer to template specificity

The aim of the initial PCR experiment was to test the annealing specificity of the primers (Ngandjio *et al.,* 2003) to the intended PCR positive control template sequence. The method described by Ngandjio *et al.,* (2003) was used to amplify *ompA* specific PCR products from two L2 isolates cultivated *in vitro.*

Two different L2 DNA samples, L2/Amsterdam/Proctitis (1.6ng/µL) and L2/434/Bu/I (1.6ng/µL), were chosen to test the CT1 and CT5 primer-to-template specificity.

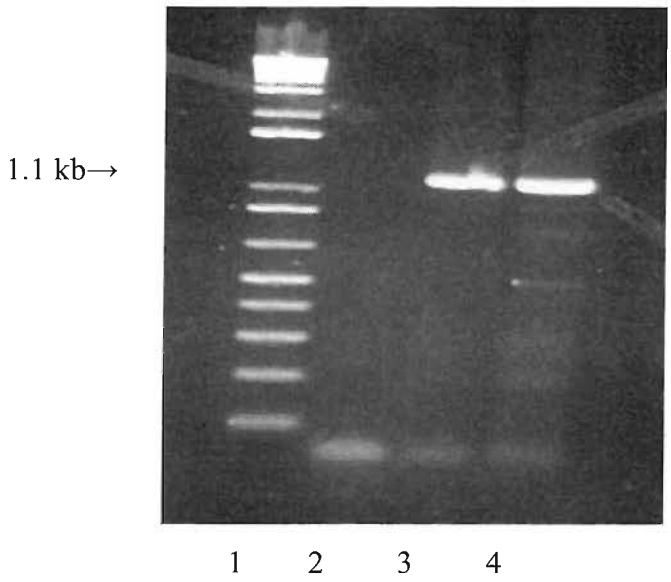


Figure 3.4: Primer-to-template specificity testing, detection of *ompA* in DNA from two L2 isolates using CT1 and CT5 primers

Lane 1: 1kb plus ladder, Lane 2: No template PCR control, Lane 3: L2/434/Bu/I (1.6ng/µL), Lane 4: L2/Amsterdam/Proctitis (1.6ng/µL).

OmpA specific products are clearly visible in lanes 3 and 4 of figure 3.4 at 1.1kb. Although other minor non specific products are visible in lane 4 the majority product is at 1.1kb. This meant that the CT1 and CT5 were specific to the intended positive PCR control template DNA.

3.2 CT1 and CT5 primer optimisation

The minor non specific banding seen in Lane 4 of figure 3.4 prompted investigation into the optimisation of the DNA amplification parameters in the PCR method described by Ngandjio *et al.,* (2003). During this optimisation a number of factors were changed. Changes included alterations in reaction components such as thermostable DNA polymerases (figures 3.5 and 3.6), investigation into the final concentration of $MgCl_2$ to optimise the annealing of the primers to the template DNA (figure 3.7), variations in the concentration of template DNA in the reaction (figures 3.8-3.12) and variations in primer annealing temperatures during the thermocycling stage of the PCR (figure 3.13).

3.2.1 Thermostable DNA polymerase optimisation

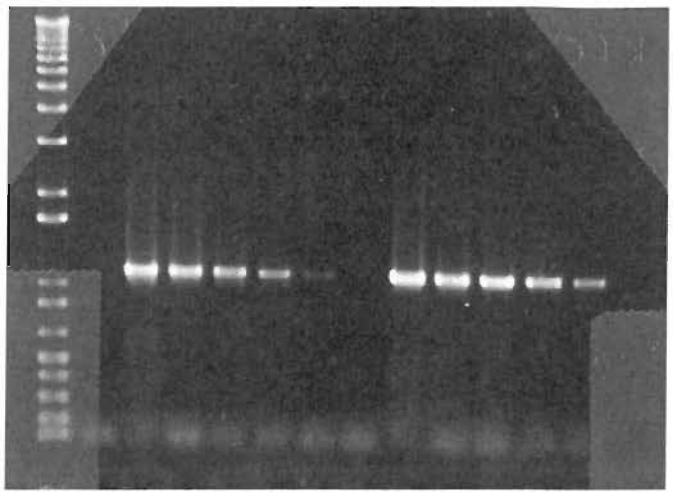


Figure 3.5: Evaluation of thermostable DNA polymerases (Invitrogen and Bio-X-Act)

In lanes 2-7 'Invitrogen' thermostable DNA polymerase was used in the PCR reactions. In lanes 8-13: Bio-X-Act short thermostable DNA polymerase was used in the reactions. L2/Amsterdam/Proctitis DNA was used as template for all reactions. The concentrations were as follows: 1.6ng/µL (lanes 3 and 9), 0.16ng/µL (lanes 4 and 10). 16pg/µL (lanes 5 and 11) 1.6pg/µL (lanes 6 and 12) and 0.16pg/µL (lanes 7 and 13) Lanes 2 and 8 are no template PCR controls.

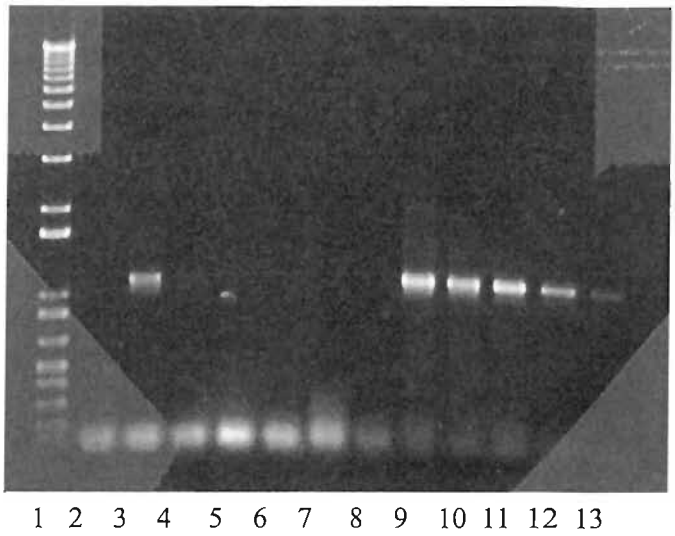


Figure 3.6: Evaluation of thermostable DNA polymerases (Promega and Red Taq)

In lanes 2-7 'Promega' thermo stable DNA polymerase was used in the reactions. In lanes 8-13: Red Taq thermostable DNA polymerase was used in the reactions. L2/Amsterdam/Proctitis DNA was used as template for all reactions. The concentrations were as follows: 1.6ng/µL (lanes 3 and 9), 0.16ng/µL (lanes 4 and 10). 16pg/µL (lanes 5 and 11) 1.6pg/µL (lanes 6 and 12) and 0.16pg/µL (lanes 7 and 13) Lanes 2 and 8 are no template PCR controls.

The results from figure 3.5 and 3.6 showed that the Bio-X-Act short thermostable polymerase had the highest sensitivity in the PCR reactions of the thermostable polymerases tested. This is shown by the clear *ompA* PCR product seen at 1.1kb even at the lowest concentration (0.16pg/µL) of L2/Amsterdam/Proctitis DNA template when compared with the other thermostable polymerases. Therefore it was decided to use this thermostable polymerase for future PCR *ompA* amplifications

3.2.2 $MgCl_2$ concentration optimisation

To optimise the *ompA* specific PCR reaction the concentration of $MgCl_2$ was varied. It was hoped that a change in $MgCl_2$ concentration may increase the sensitivity (i.e. detection limit) of the assay.

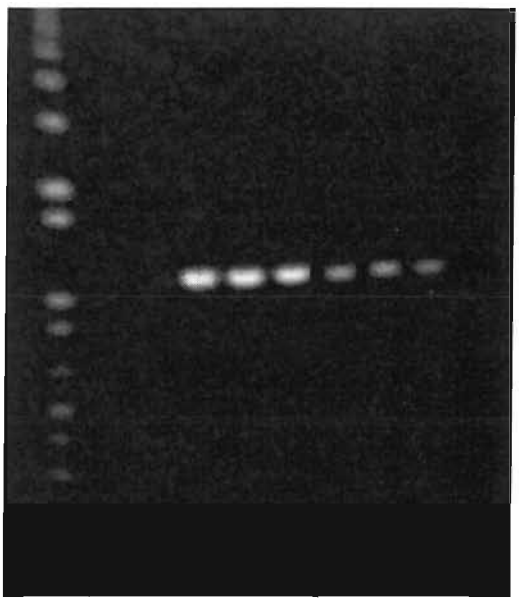


Figure 3.7: Optimisation of the *ompA* specific PCR $MgCl_2$ concentration using CT1 and CT5 primers and L2/Amsterdam/Proctitis (1.6ng/µL) as DNA template

Lane 1: 1kb plus ladder, Lane 2: No template PCR control, Lane 3: 0mM $MgCl_2$, Lane 4: 1.25mM $MgCl_2$, Lane 5: 2.5mM $MgCl_2$, Lane 6: 3.75mM $MgCl_2$, Lane 7: 5mM $MgCl_2$ Lane 8: 6.25mM $MgCl_2$,Lane 9: 7.5mM $MgCl_2$

From Figure 3.7, the highest concentration of *ompA* specific PCR product was seen in lanes 4, 5 and 6. It was decided there was an optimal range for the concentration of $MgCl_2$ in the PCR between 1.25mM and 3.75mM. Based on this range a concentration of 3.5mM $MgCl_2$ was chosen for future reactions. It was thought that probable DNA template concentration variations in the samples received from Jersey might affect the annealing of the primers. The slight excess of $MgCl_2$ in the reaction was hypothesised to increase the efficiency of

the thermostable DNA polymerase during amplification by counteracting the overall negative repelling charge of the DNA and primers. This potentially could yield more PCR products to be genotyped via sequencing.

3.3 BD Probe Tech samples selected to trial optimised PCR conditions

Once DNA amplification conditions had been optimised, using DNA template extracted from laboratory LGV strains, the new protocol was tested on the six selected samples received from Jersey in the batch dated 12.10.05. These samples were selected as the received volume was at least 0.5mL. It was envisaged there would be a need for multiple use of each sample during the developmental stages of the project. The samples were also selected on the basis of the MOTA scores of the BD Probe Tech diagnostic assay. It was hypothesised that although the MOTA score did not directly reflect the quantity of chlamydiae in a sample it might give an indication of the potential amount of target available for the PCR reaction. It was hypothesised that high MOTA scores might be more likely to produce positive *ompA* specific PCR results. In addition if the MOTA score could be linked to the quantity of potential *ompA* specific PCR product then it was also hypothesised to be feasible to determine the limits of detection of the PCR from the MOTA scores of the BD Probe Tech diagnostic assay. Based on these hypotheses, the six samples chosen represented the extremes of the MOTA score range available. The sample data is shown in table 3.5.

Table 3.5 Details of the six selected samples chosen to trial the optimised *ompA* specific PCR amplification conditions.

Sample Number	MOTA score
11	21872
14	2040
18	59534
24	2310
31	26276
50	60292

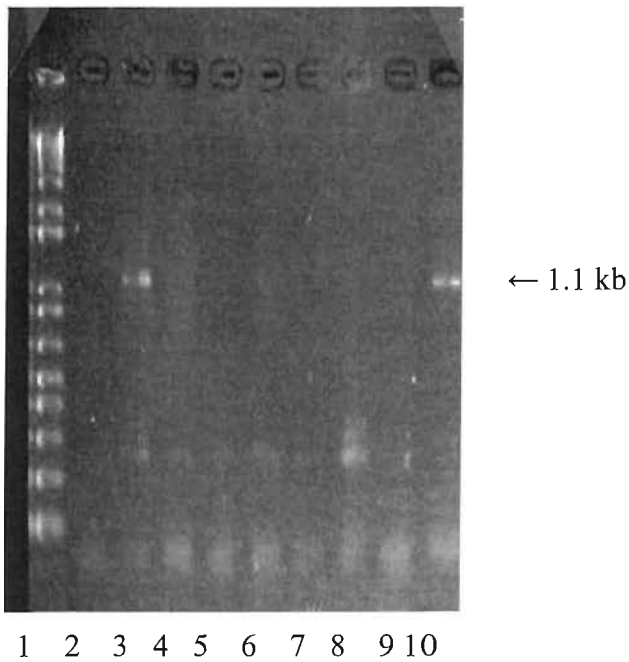


Figure 3.8 Six selected samples chosen to trial the optimised *OmpA* specific PCR amplification conditions

Lane 1: 1kb plus ladder, Lane 2: no template PCR control, Lane 3: Sample 11, Lane 4: Sample 14, Lane 5: Sample 18, Lane 6: Sample 24, Lane 7: Sample 31: Lane 8: Sample 50, Lane 9 no template PCR control, Lane 10: L2/Amsterdam/Proctitis (1.6ng/µL) positive control.

The amplification of the six samples resulted in only a single *ompA* specific PCR product. This was from sample 11, a sample with a mid range MOTA score of those selected. Within the range of selected samples, it was thought that 11 and 31 shared close mid range MOTA scores. Due to this similarity in scores, but the difference in *ompA* positivity via PCR of these samples, it was envisaged that it might not be possible to get all of the samples to produce an *ompA* specific positive result via the optimised PCR.

3.4 Initial *ompA* screening of ten clinical samples from Jersey

Initial hopes were that the optimised DNA amplification reaction conditions would give *ompA* specific positivity in 100% of the BD Probe Tech *C.trachomatis* positive samples received from Jersey. However Figure 3.8 showed that this was a very optimistic target. The problems in obtaining PCR products from all of the samples were attributed to the quality and initial quantity of the template in the BD Probe Tech *C.trachomatis* positive samples. There seemed little benefit from quantifying the DNA template in the BD Probe Tech *C.trachomatis* positive samples upon receipt, due to their low volume. In addition, even with the amplification of the template DNA, positive visualisation of the PCR product was not apparent. However, the first 10 samples received from Jersey, in the batch dated 12.10.05 were tested for *ompA* positivity in an attempt to gauge a potential positivity percentage for future reference. These samples came from a batch of 50 samples archived in the Southampton laboratory.

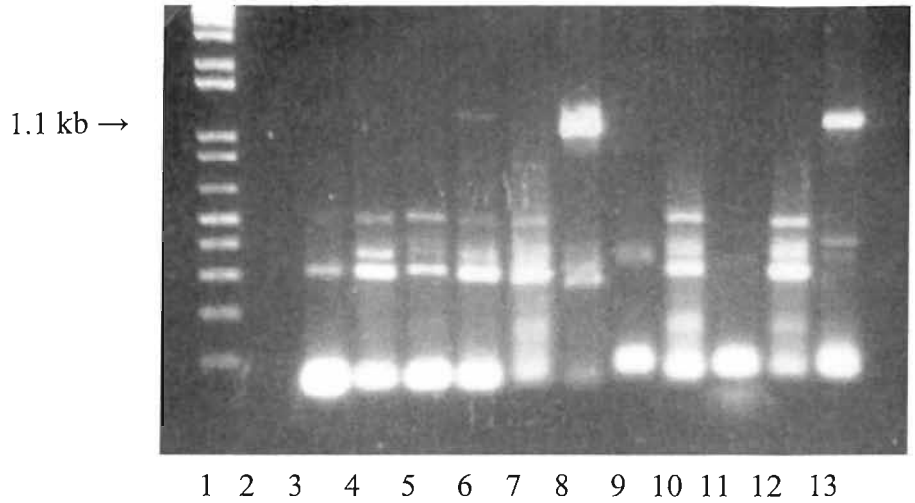


Figure 3.9: Initial *ompA* screen of the first 10 clinical samples received from Jersey

Lane 1: 1kb plus Ladder, Lane 2: No template PCR control, Lanes 3-12: Samples 1-10 of Jersey sample batch 12.10.05 Lane 13: L2/Amsterdam/Proctitis (1.6ng/μL) positive control

From Figure 3.9 clear *ompA* positive banding can be seen in lane 6 corresponding to sample 4 and lane 8 corresponding to sample 6. It was envisaged that the quantity of PCR product visualised in lane 6 by sample 4 would be insufficient to be able to recover from the agarose gel and further genotype however; both *ompA* specific bands were gel extracted using the method outlined in section 2.17.1 and archived for future genotyping via sequencing.

From the results shown in figure 3.9 it was conservatively estimated that approximately 10% of the samples received from Jersey would firstly be positive for the *ompA* gene using the proposed DNA amplification method and secondly, would provide the quantity of amplified PCR product (40ng minimum) that could be gel extracted and genotyped via sequencing.

During 2005, the first year of the project, 72 BD Probe Tech *C.trachomatis* positive samples were received from Jersey. All were selectively amplified using the CT1 and CT5 primers for the *ompA* gene. Of the 72 samples received from Jersey in 2005 ten gave *ompA* positive PCR results.

It was noted that although all the BD Probe Tech *C.trachomatis* positive samples from Jersey were *C.trachomatis* positive in that commercial test, amplification of the *ompA* gene from these templates was not 100% successful. This could be attributed to PCR inhibitors such as blood and haem in the samples. It was also thought that there may have been some degradation of the DNA target in the sample during transit or storage. It was thought that the dilution of the samples received from Jersey would decrease the concentration of the possible PCR inhibitors and allow for the successful PCR amplification of the required *ompA* product from more of the samples.

3.5 Dilution of original samples to potentially dilute PCR inhibitors

The BD Probe Tech *C.trachomatis* positive samples were initially tested from undiluted material and also as dilutions of 10 and 100 fold of the original sample using UHQ water, to reduce the concentration of any inhibitors of the DNA amplification reaction.

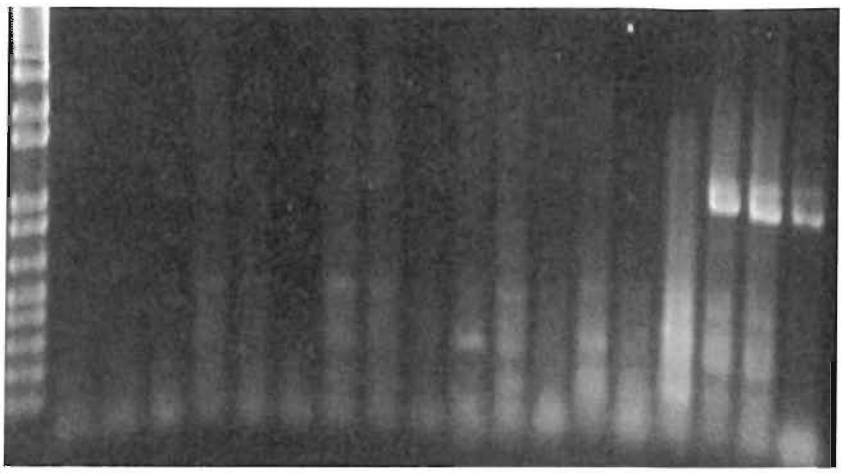


Figure 3.10: Dilution of template DNA entering the *ompA* PCR

Lane 1: 1kb plus ladder, Lanes 2,5,8,11,14, and 17: Samples 1-6 of Jersey sample batch 12.10.05, Lanes 3,6,9,12,15 and 18: 1:10 dilution of samples 1-6 from Jersey sample batch 12.10.05, Lanes 4,7,10,13,16,and 19: 1:100 dilution of samples 1-6 from Jersey sample batch 12.10.05.

It was also hoped that the faint positive *ompA* band of sample 4 (seen in lane 6) of figure 3.10 would produce a greater *ompA* specific PCR product yield if PCR inhibitors were diluted from the reaction. However, from figure 3.10 this proved to be unfulfilled and no *ompA* specific product was seen in either the original or the ten fold or one hundred fold dilutions of the neat sample. The cause of this could be the result of repeat freeze thawing of the samples causing template degradation. However, it could also be attributed to the uneven mixing of the sample after thawing. To rule this factor out in later experiments, all thawed samples were vortexed for 10 seconds prior to use in DNA amplification reactions.

3.6 Adjustments to the reaction volume to potentially reduce the effects of PCR inhibitors without altering the amount of DNA template entering the reaction

By diluting the Jersey samples 10 fold and 100 fold of the neat sample (to decrease the concentration of potential inhibitors) there was also a dilution of the potential template for the DNA amplification reaction. This was thought to be a problem, as due to the nature of the samples collected from Jersey, the quantity of template originating in the sample was known to be of varying quantity due to the variations seen in the six samples initially selected to trial the *ompA* specific PCR and, the amplification of the first 10 samples from the Jersey sample batch dated 12.10.05. Genotyping of each positive *ompA* specific product would require a minimum of 40ng of quality amplified PCR product. Diluting the DNA template entering the reaction with water to remove inhibitors was thought to be inefficient as there was the possibility of exceeding the limits of detection for visualising the PCR product on agarose gels.

To combat this problem, it was proposed that by adjusting the volume of the DNA amplification reaction but keeping reaction mix concentrations constant it was possible to adjust the concentration of the samples received from Jersey. In doing so the concentration of possible potential inhibitors of the PCR amplification would be reduced however, the amount of template being added to the reaction would remain constant. To test this, variations in the overall volume of the reaction were completed.

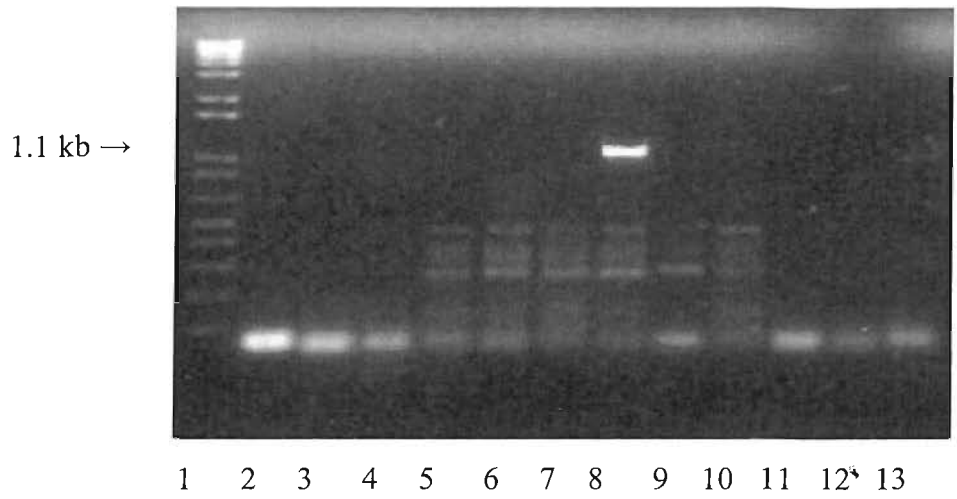


Figure 3.11: Comparative evaluation of reaction volume

1µL of template was added to 20µL of reaction master mix

Lane 1: 1kb plus ladder, Lane 2: No template PCR control, Lane 3-12: Samples 1-10 of Jersey sample batch 12.10.05, Lane 13: No template PCR control

Figure 3.11 shows the original component concentration and volume of the DNA amplification reaction. All comparative observations were made in relation to figure 3.11.

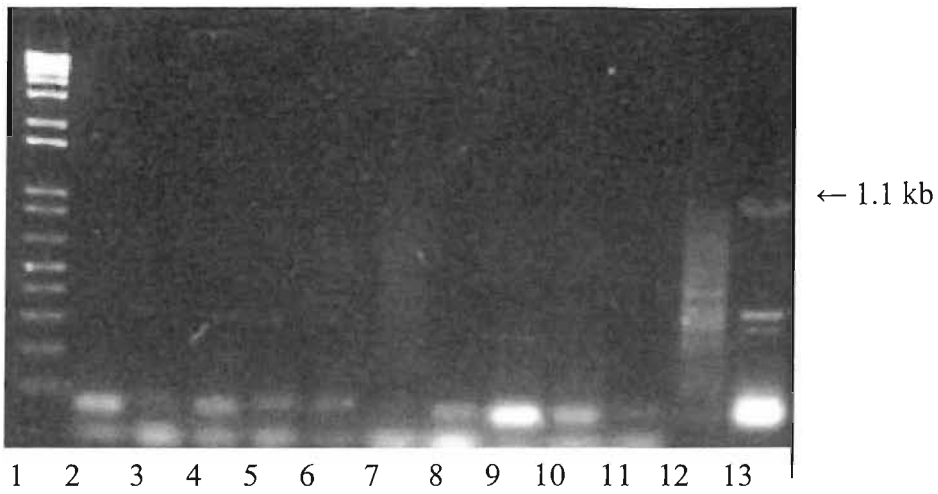


Figure 3.12: Comparative evaluation of reaction volume

1µL of template was added to 50µL of reaction master mix

Lane 1: 1kb plus ladder, Lane 2-11: Samples 1-10 of Jersey sample batch 12.10.05, Lane 12: No template PCR control, Lane 13: L2/Amsterdam/Proctitis (1.6ng/µL) positive control

Clearly increasing the volume of the DNA amplification reaction altered the concentration of the components within the reaction. However, it was noted that this was not an approach that would increase the efficiency or the sensitivity of the PCR reaction. The *ompA* positive PCR product routinely seen from sample six of the Jersey sample batch 12.10.05 is not present in this figure and clear primer dimer banding can be seen at approximately 80bp in all the samples tested. This shows that the primers have not bound to the template as intended but rather to each other and so amplification of the intended *ompA* product was not achieved.

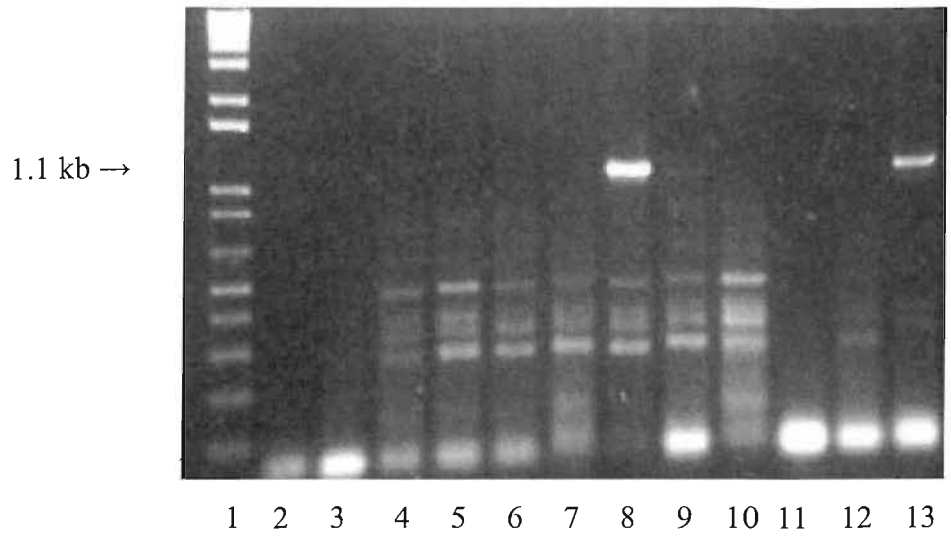


Figure 3.13: Comparative evaluation of reaction volume

2µL of template was added to 50µL of reaction master mix

Lane 1: 1kb plus ladder, Lane 2: No template PCR control, Lane 3-12: Samples 1-10 of Jersey sample batch 12.10.05, Lane 13: L2/Amsterdam/Proctitis (1.6ng/µL) positive control.

Figure 3.13 shows that a less extreme dilution, with master mix, in the samples received from Jersey gave similar results to the original 20µL reaction. The master mix used in the reactions shown in Figures 3.11, 3.12 and 3.13 was universally mixed and aliquoted accordingly to each amplification reaction and so variations in the quantities of reaction chemical components or the possibility of missing out a component of a reaction due to researcher error is not a factor. Therefore the results seen must be a true reflection of the effects of adding more DNA template.

Because altering the reaction volume and amount of DNA template in the PCR reaction had had no positive effect on the resulting PCR *ompA* specific product output, attention was directed to the annealing temperature specificity of the PCR.

3.7 Comparative evaluation of primer annealing temperature

It had been noted that primer annealing was not exclusive and produced multiple banding when amplicons were visualised on a 1% agarose gel. This was an increasing problem as the amplicons visualised in the non specific bands were thought to be increasing in concentration over time. To try and rectify this problem an alteration was made to the annealing temperature of the DNA amplification reaction. It was envisaged that by increasing the annealing temperature of the primers there would be an increase in the specificity of primer annealing to the template and also the PCR *ompA* specific PCR product yield. The annealing temperature was changed from 57^0C to 60^0C.

The altered amplification conditions described were as follows

Temperature	Time
95^0C	5 minutes
35 Cycles of	
95^0C	1 minute
60^0C	1 minute
72^0C	1 minute
Final extension step	
72^0C	10 minutes

Amplicons produced were electrophoresed on the same 1% agarose gel for comparative purposes.

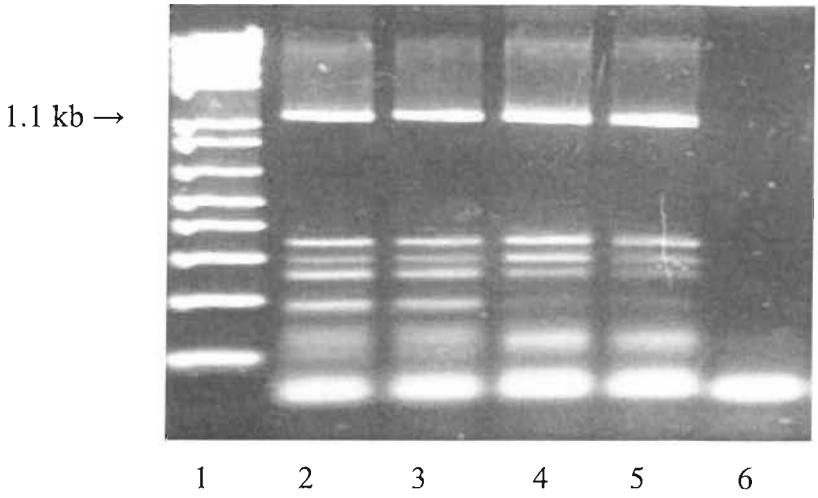


Figure 3.14: Comparative evaluation of primer annealing temperature. Lane 1: 1 kb plus ladder, Lane 2: L2/Amsterdam/Proctitis (1.6ng/µL) at a 57^{0}C annealing temperature, (positive control) Lane 3: L2/Amsterdam/Proctitis (0.16ng/µL) at a 57^{0}C annealing temperature, Lane 4: L2/Amsterdam/Proctitis (1.6ng/µL) at a 60^{0}C annealing temperature, Lane 5: L2/Amsterdam/Proctitis (0.16ng/µL) at a 60^{0}C annealing temperature, Lane 6: No template PCR control.

Figure 3.14 shows that increasing the annealing temperature to 60^{0}C did reduce some of the non specific banding seen from the L2/Amsterdam/Proctitis template. However, it did not stop all of the non specific amplicons being made during the reaction as was hoped. In addition, there was no significant increase in the amount of *ompA* specific PCR product seen when the annealing temperature was increased to 60^{0}C and so the original PCR amplification conditions outlined in the methods described by Ngandjio *et al.,* (2003) were reverted to.

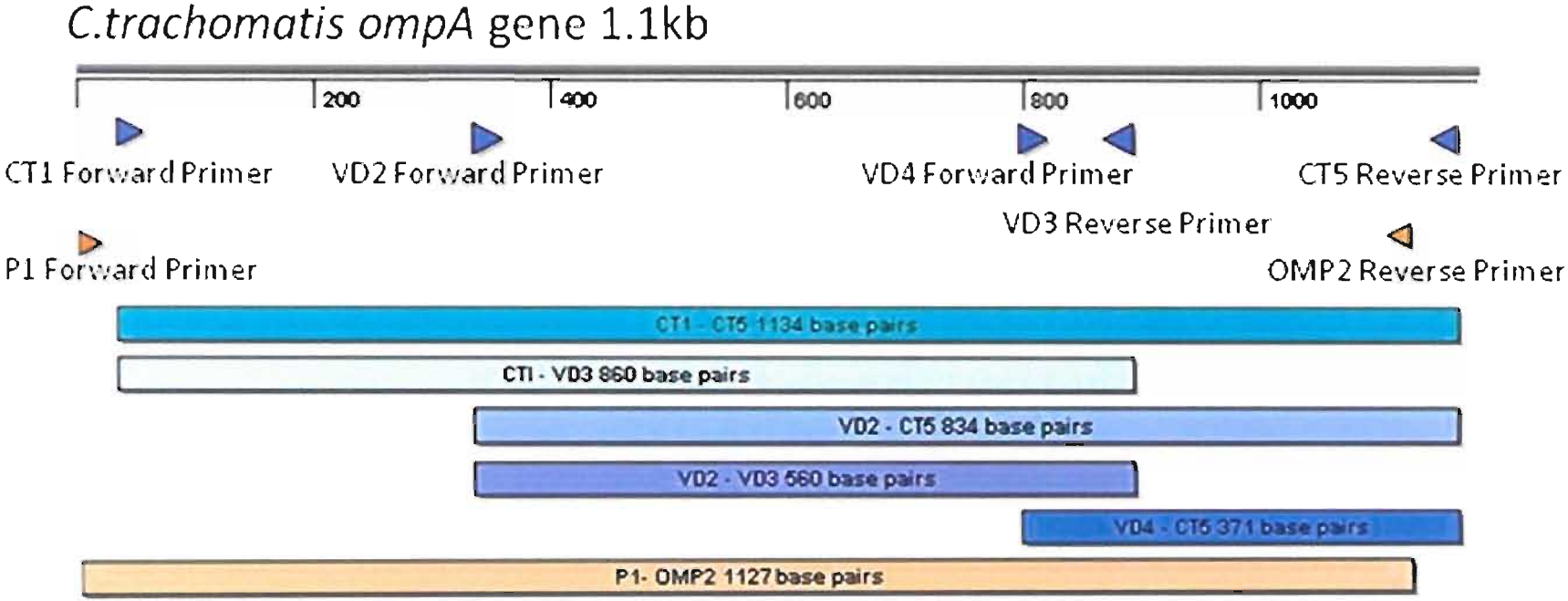
The increase in the production of non specific products was attributed to sample degradation due to multiple uses over time as each time the sample was used it was thawed from frozen a process known to fragment DNA. In addition the original design and specificity of the primers to the intended template was again considered.

3.8 Amplification of specific *ompA* variable domains

Because of the limited success in amplifying specific *ompA* PCR products from the samples received from Jersey and the risk of sample degradation over time, it was thought that the quality of the template in many samples could be too poor to amplify the 1.1kb fragment of *ompA* in its entirety. This prompted investigation into a range of primers and primer combinations that were designed to amplify individual variable domains of *ompA.* It was thought that the efficiency of the reaction and specificity of the primer annealing could be increased by amplifying smaller fragments as specific domains may have been sufficient to allow genotyping by sequencing. The following primers were tested to evaluate template annealing specificity, and consisted of the internal sequencing primers of (Ngandjio *et al.,* 2003) and the primers of an alternative *ompA* study by Jurstrand *et al.,* (2001)

Each primer combination was mapped on to the L2/434/Bu *ompA* sequence using the Lasergene 6 computer package. This was to give an indication of the size of the specific *ompA* variable domain PCR products required and as a point of reference when characterising PCR products from electrophoresis gels.

Figure 3.15: Primer locations and PCR product size



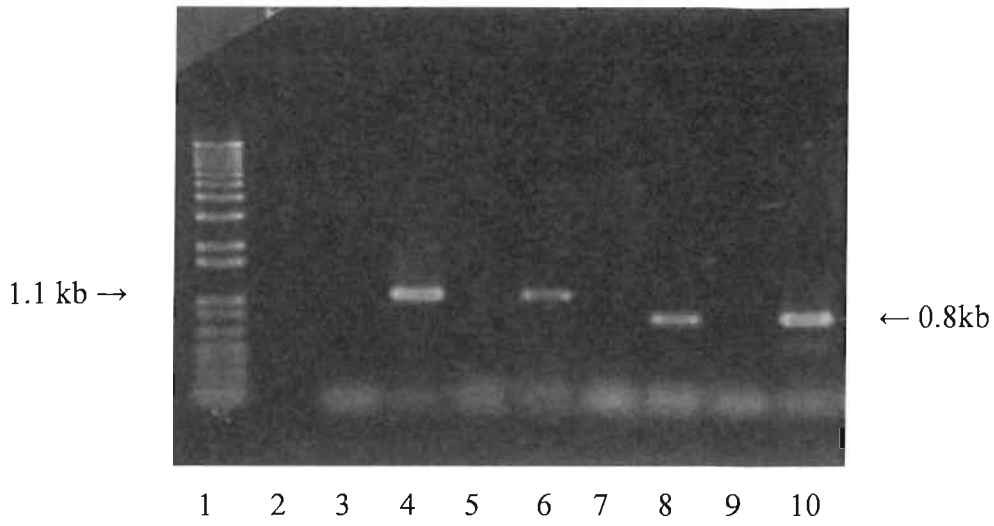


Figure 3.16: Comparative evaluation of primer combinations outlined in the methods described by Jurstrand *et al.,* (2001)[1]; Ngandjio *et al.,* (2003)[2].

Lane 1: 1kb plus ladder, Lane 2: No primer PCR control, Lane 3: No template PCR control of P1-OMP2[1] primer master mix , Lane 4: P1-OMP2[1] *ompA* PCR, Lane 5: No template PCR control CT1-CT5[2] primers, Lane 6: CTI-CT5[2] amplicon PCR, Lane 7 No template PCR control VD2-CT5[2] primers, Lane 8: VD2-CT5[2] amplicon PCR Lane 9 No template PCR control CT1-VD3[2] primers, Lane 10: CT1-VD3[2] amplicon PCR, DNA template in all reactions was L2/Amsterdam/Proctitis (1.6ng/µL)

Figure 3.16 showed that the L2/Amsterdam/Proctitis positive control DNA could be amplified with any of the combinations of primers outlined in the methods described by Jurstrand *et al.,* (2001); or Ngandjio *et al.,* (2003). However, Figure 3.16 lane 4 shows that the greatest PCR product concentration was seen with the P1-Omp2 primer combination. A product of identical intensity was noted in lane 10 of figure 3.16, however, as this product didn't encompass the whole of the *ompA* gene it was preferable to use a combination of primers that did select and amplify the whole gene.

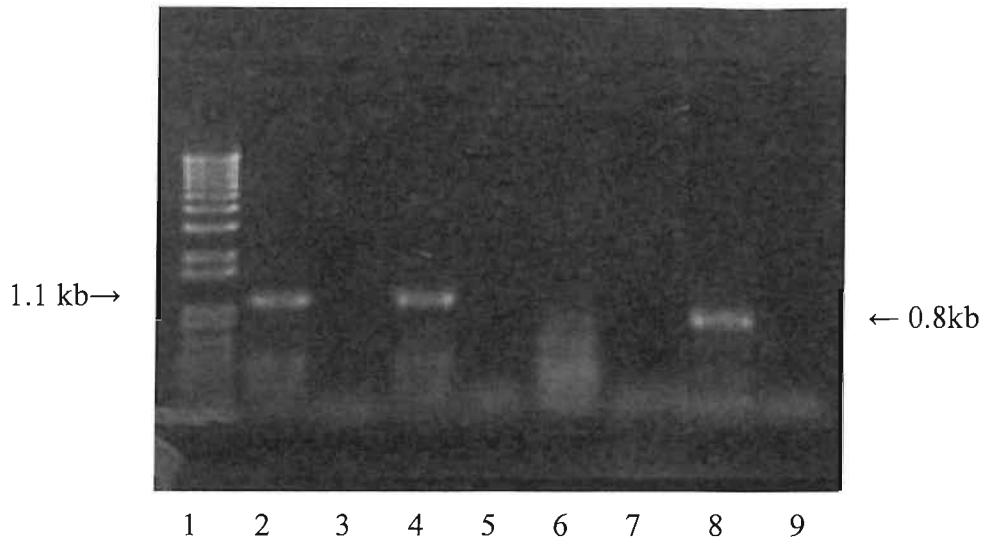


Figure 3.17: Comparative evaluation of primer combinations outlined in the methods described by Jurstrand *et al.*, (2001)[1]; Ngandjio *et al.*, (2003)[2].

Lane 1: 1kb plus ladder, Lane 2: P1-OMP2[1] *ompA* PCR, Lane 3: No template PCR control of P1-OMP2[1] primer master mix, Lane 4: CTI-CT5[2] amplicon PCR, Lane 5: No template PCR control CT1-CT5[2] primers, Lane 6: VD2-CT5[2] amplicon PCR, Lane 7 No template PCR control VD2-CT5[2] primers, Lane 8: CT1-VD3[2] amplicon PCR Lane 9 No template PCR control CT1-VD3[2] primers. DNA template in all reactions was Sample 6 from Jersey sample batch 12.10.05.

Figure 3.17 showed that clinical sample 6 could either be amplified with the P1-Omp2[1], CT1-CT5[2] or the CT1-VD3[2] combination of primers, to provide a PCR product that could be genotyped via sequencing. However, the CT1-VD3 combination of primers did not encompass the whole *ompA* gene and so was rejected. It was decided that the quality or quantity of the PCR products was not significantly greater with one combination of primers or the other when considering either the P1- Omp2[1] or CT1-CT5[2] combinations of primers and so the original combination of CT1-CT5 was used for future work. However, the quality of the resulting sequencing data of this amplicon was still very poor and required multiple sequencing attempts to obtain quality data.

Using smaller volumes of the *ompA* specific PCR product of a higher quality was thought to be a more efficient approach than sequencing multiple times those that were producible, this prompted investigation into the clean up of the original DNA samples received from Jersey.

3.9 DNA template clean up methods

Three DNA template purification methods were tested in an attempt to increase the template quality entering PCR reactions. QIAamp spin columns (Qiagen), the use of a Chelex resin clean up method (Jurstrand et al, 2001) and the automated MagNA Pure LC system (Roche) were evaluated. It was thought that by purifying the template entering the initial *ompA* specific DNA amplification reaction, subsequent amplicons would be of a greater quality and this may give rise to better sequencing results. In addition, not all of the samples proved positive by the BD Probe Tech resulted in a positive using the *ompA* specific DNA amplification method and it was hoped that by further extracting and purifying the DNA from samples, inhibitors of the PCR reaction would be removed and thus *ompA* selective DNA amplification would be improved during the DNA amplification reaction.

Five BD Probe Tech *C.trachomatis* positive samples received from Jersey were selected, the details of these samples are shown in table 3.6

3.10 Chlamydia DNA template purification

Quality template (i.e. with suitable intact DNA to allow PCR of the desired products and as free as possible of reaction inhibitors) is required for PCR as impurities may impede specific amplicon amplification. This means that quality DNA template is required for amplifications to get the best sensitivity, avoid contamination and inhibition which will undoubtedly affect the downstream applications of the amplification products such as sequencing. To try and increase the yield of the required specific amplified products from the PCR reactions three template purification methods were evaluated to see if better quality sequencing results could be achieved by adding purer DNA template to the original *ompA* amplicon PCR reaction. A range of samples was chosen from the same sample set on the basis that during an initial *ompA* screen with the CT1 and CT5 primers they had a weak positive *ompA* signal or no signal at all and were though to contain PCR inhibitors. One sample deemed to be negative was spiked with 10µL of of L2/Amsterdam/Proctitis 1.6ng/µL DNA.

Table 3.6: Details of five samples received from Jersey selected to test three DNA template clean up methods

Sample Number	*ompA* positive using the method described by Ngandjio *et al.*,2003	*MOTA Score*
16	Very faint positive	41653
26	no	33827
45	no	35870
46	no	29893
47	no	20548

All samples were received in the batch from Jersey dated 14.02.06.

It was noted that although all the BD Probe Tech screened samples received from Jersey were confirmed *C.trachomatis* positive via the BD Probe Tech diagnostic assay, amplification of the *ompA* gene from these templates using primer pairs CT1-CT5, VD2-CT5 CT1-VD3 and PI-OMP2 had not previously been successful. To ensure a positive control for the clean up methods sample 45 was spiked with 10µL of L2/Amsterdam/Proctitis 1.6ng/µL DNA.

For each of these DNA clean up methods, 100µL of each sample was diluted in 2mL water to give a comparable volume and concentration of 'cleaned up' DNA. The final elution for each of the samples, from each of the methods was 50µL. 1µL of this eluate, for each sample from each method, was amplified using the previously described conditions, the results are summarised in table 3.7.

Table 3.7: Detection of *ompA* in DNA from three DNA template clean up methods using CT1 and CT5 primers

Sample numbers	16	26	45 (spiked)	46	47
Jersey sample neat	+(Very faint)	-	+	-	-
MagNA Pure (Roche)	-	-	+	-	-
Chelex ion exchange resin	-	-	+	-	-
QIAamp columns (Qiagen)	-	-	+	-	-

+ = OmpA positive PCR product - = OmpA negative PCR product

To test whether the quality of the PCR products had been improved by the clean up procedures the CT1-CT5 amplicons of sample 45 from Jersey batch 14.02.06 were subjected to DNA sequence analysis

3.11 DNA sequencing of *ompA* positive samples after clean up of template DNA entering the PCR reaction

Sequencing data suggested that the 'clean up' methods tested gave no significant difference in the quality or quantity of the template entering the sequencing reactions or to the sequencing results of PCR products. It was finally concluded that the required *ompA* PCR products could not be amplified from all of the clinical samples received from Jersey and the 'clean up' methods gave no additional benefits.

3.12 Pooling of identical PCR products to give a more concentrated PCR product template for DNA sequencing

The volume of template used in the sequencing reactions was <5µL. However, the quality of sequencing results was variable, sequencing reactions frequently had to be repeated and this rapidly used up PCR amplicons creating reproducibility problems. To overcome this it was necessary to have an increased quantity of product from the initial *ompA* amplification as some of this product was lost during the purification of nucleic acids by gel extraction. Therefore it was decided to scale up *ompA* PCR reactions to the equivalent of three 1:20 PCR reactions per sample (3:60) and pool the gel extracted *ompA* PCR product to provide enough identical template DNA to go into the initial sequencing reaction and any necessary repeats.

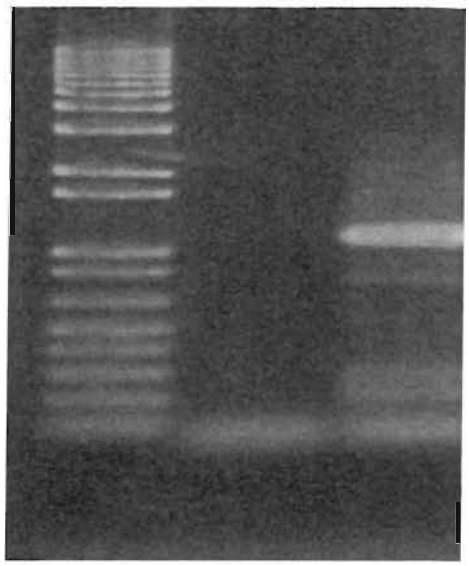


Figure 3.18: Recovery gel to demonstrate the gel extraction and pooling of PCR reactions from samples received from Jersey

Lane 1: 1kb plus ladder, Lane 2 a single reaction showing 'lost' band during gel extraction. Lane 3: DNA amplification reaction x3, pooled and then gel extracted.

Lane 3 of figure 3.18 clearly shows an *ompA* specific product at 1.1kb; this is the result of pooling the amplification reaction product and therefore a better recovery of the PCR product. Lane 2 shows the total loss of the *ompA* specific PCR product, during recovery, through gel extraction. The 'lost' PCR product was a single 1:20 PCR reaction identical to those that have been pooled in lane 3.

In addition to increasing the amount of PCR product being loaded on to the visualisation gel, a different method of gel staining was tested to see if it would be more sensitive in detecting PCR products than the ethidium bromide method already used. It was hoped that if the method proved more sensitive less PCR product would need to be loaded onto the gel and therefore another PCR clean up method might be suitable to ultimately give a sequencing template rather than have to use gel extraction procedures.

3.13 Comparisons in agarose gel staining methods

During the development of gel extraction methods two different gel staining methods were tested, ethidium bromide and SYBR green. 2.5ng/mL ethidium bromide (Fisher) was used for staining. After the agarose gel had been electrophoresed it was submerged fully in the ethidium bromide solution. The gel was stained for 1 hour. Excess ethidium bromide was washed away with water. In contrast the samples shown in figure 3.20 had 1µL of SYBR green added to the sample during gel loading and this was electrophoresed with the sample through the gel. Gel extractions were completed of the *ompA* specific positive PCR products from each of the stained gels as described in section 2.15.2. To determine whether, the gel extracted PCR products recovered from 1% agarose gels, were suitable for sequencing they were diluted 10 fold and 100 fold with UHQ water and re-amplified with the same primers and the optimised *ompA* PCR described previously.

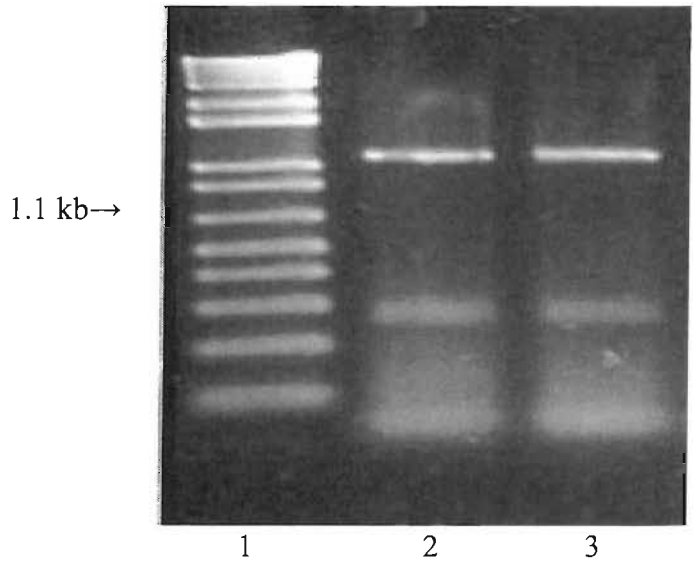


Figure 3.19: Comparative evaluation of staining methods (ethidium bromide)

Samples were stained with ethidium bromide Lane 1: 1kb plus ladder, Lane 2: L2/Amsterdam/Proctitis (secondary amplification of CT1-CT5 PCR product) Lane 3: L2/Amsterdam/Proctitis (primary amplification of original template)

Figure 3.19 shows that CT1-CT5 PCR amplified material can successfully be recovered from an electrophoresis gel stained with ethidium bromide as outlined in section 2.14.1 of the methods. The samples were excised from the gel, gel extracted and re-amplified under identical conditions. This suggested that these primers may also have been suitable for sequencing of *ompA* specific amplicons.

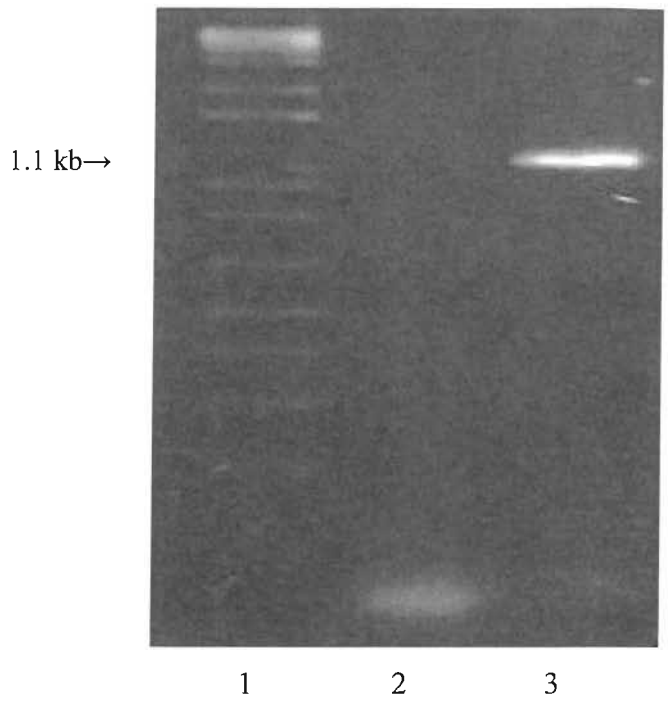


Figure 3.20: *OmpA* specific PCR amplicons stained with SYBR Green

Samples were stained with SYBR Green. Lane 1: 1kb plus ladder, Lane 2: L2/Amsterdam/Proctitis (secondary amplification of CT1-CT5 PCR product) Lane 3: L2/Amsterdam/Proctitis (primary amplification of original template)

Figure 3.20 lane 3 shows that *ompA* specific PCR positive products can be visualised when stained with SYBR green.

A comparative analysis of the gel staining process was completed where it was found that *ompA* specific PCR product extracted from a gel stained with SYBR Green could not be amplified further when used as template for a subsequent DNA amplification reaction. This is shown by the *ompA* specific PCR product present in lane 3 of figure 3.20 and the absence of the *ompA* specific PCR product in lane 2 of figure 3.20

It was decided that all future gels should be stained with ethidium bromide and speculated that the SYBR green was intercalated so tightly with the DNA that when the DNA was eluted from the gel extraction spin column that the SYBR green was being eluted with it and masking the nucleotides in any further PCR so that the DNA could not be amplified further.

4.0 Sequencing development

Clinical samples received from Jersey were successfully amplified by PCR to give *ompA* amplicons which were genotyped via sequencing and comparatively analysed with the published sequences of Yuan *et al.,* (1989). Initially *ompA* specific products from L2/Amsterdam/Proctitis and sample 6 of the 12.10.05 sample batch received from Jersey were used to develop sequencing using the internal primer VD4 and the external primer CT5 (see figure 3.16).

After the initial sequencing of two PCR products, the sequencing reaction technique was noted as being extremely sensitive to variation. A result of this was extreme differences in data quality ranging from 'unreadable' chromatograms, to data of a quality that could determine single nucleotide differences as was the case in the data used to confirm L2/Amsterdam/Proctitis as an LGV2b strain. It was evident that the sequencing would require patience and practice to produce the consistent quality data required to genotype the *C.trachomatis* positive samples received from Jersey. A consistent sequencing output would allow the comparative analysis of the data with the published sequences (Yuan *et al.,* 1989).

4.1 Comparative analysis of *ompA* sequences

4.1.1 Comparative analysis of LGV 2 and sub strains

Every *C.trachomatis ompA* gene contains a high percentage of genetic identity. However, each also has specific point mutations and other sequence variations, facilitating genotyping, and this is also reflected serologically to give a corresponding specific serovar. To fully understand the importance and significance of point mutations and the variations between genotypes they cause at a molecular level, and to practice the sequencing technique, a comparative analysis was completed of a selection of LGV strains. This comparative analysis of the LGV2 strain and possible sub-strains and the historical tracking specifically of the L2b strain was completed in anticipation of the L2b strain presenting in the UK through sexual networking with Europe. Europe had recently witnessed an unusually large increase in the activity of this strain amongst the MSM community, prompting investigation into whether it was a new strain.

A variant of the L2 strain known as L2b has been reported to be causing a string of severe proctitis cases in Europe and New York and has been associated with MSM who also have concurrent HIV infections (Spaargaren J *et al.,* 2005)

It is important to acknowledge and understand the genetic difference between variants within a serovar, such as is the case with L2 and L2b, as these differences could ultimately account for the differences in virulence, antibiotic sensitivity, manifestation and progression of the disease. This increased understanding is essential for the treatment and management of the disease both at a local level and epidemiologically within wider populations. With this in mind a paper-based, comprehensive comparative study of the published *ompA* sequences of the L2 serovars was completed, this included the L2a and L2b variants (Thomson *et al .,* 2008).

During this comparative analysis, sequencing of an 'anonymous' sample from collaboration with the Sanger centre was completed in Southampton. The sample was processed identically to the samples received from Jersey using the CT1-CT5 primers and was successfully genotyped as a strain of the lymphogranuloma venereum serovar L2 (known as L2b).

In essence, comparative analysis of the *ompA* gene, from the LGV serovar L2 and the strains within the serovar, L2a and L2b was completed. The *ompA* gene sequences of L2, L2a and L2b were thought to be extremely conserved and showed high sequence identity. There was only a single nucleotide difference between the L2/434/Bu and L2b strain at position 485. This change was from adenine to guanine respectively; this resulted in an amino acid change during translation from asparagine to serine. In addition, two inverse repeats were noted in the sequences each of ten bps in length. The first was sited at position 573-583 within the L2b sequence and corresponded to position 660-670 within the (Stephens 1987) L2/434/Bu sequence. The second inverted repeat was located at position 610-620 within the L2b sequence and corresponded to position 623-633 within the L2/434/Bu sequence. The comparative analysis of the stains in the L2 serovar also revealed a difference over the L2/434/Bu sequence itself. It is noted that two L2/434/Bu *ompA* gene sequences have been published (Hayes *et al.,* 1994;Stephens *et al.,* 1987). These two sequences were noted to have a single bp difference between them at position 471 from a cytosine to a guanine respectively; however the translated amino acid sequence was not affected.

L2a and L2b have been documented as sub-strains within the L2 serovar because if tested serologically both show enough identity to appear as L2. However they are both in a molecular sense different from the L2/434/Bu strain. Two further strains from the trachoma biovar, both documented as serovar D, were chosen to further highlight the subtle molecular differences noted by the LGV comparative analysis. In addition it was essential to note that

although strains maybe of different origins, and have point mutations within the *ompA* genetic sequence, identity can be seen in the sequence genotypically however the point mutations may not always cause a change in serovar.

To investigate this concept further, two isolates serotypically thought of as serovar D were processed identically to the LGV strains. Both samples were sequenced using the internal VD4 primer and the external CT5 primer. The data was comparatively analysed to the published sequence for each serovar (Yuan *et al.,* 1989)

4.2 Comparative analysis of Nills Lycke and IOL1883 strains

Both of the strains from the genetic sequence were confirmed to be genotypically identical to Serovar D, however there were some point mutations in the IOL1883 sequence these are highlighted in blue (Figure 4.1). The nucleotides highlighted in yellow are those thought to be significant of genotyping a D sequence. Only one of the point mutations of the eight highlighted corresponds with those necessary for the determination of a D sequence and this was thought to be negligible as all of the other significant points are consistent with the D sequence

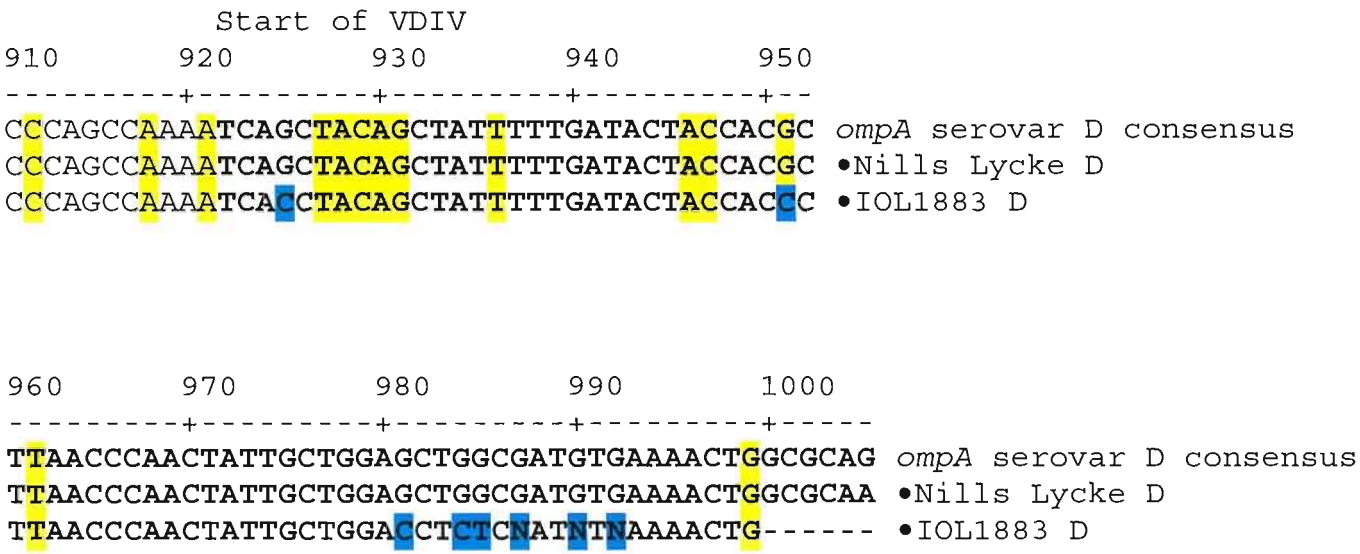


Figure 4.1: A comparative alignment of variable domain 4, serovar D.

The Yellow highlights are characteristics typical of a serovar D isolate.

The Blue highlights are miss-matches within the aligned nucleotide sequences.

4.3 Screening of samples received from Jersey

From May 2005–May 2006 192 *C.trachomatis* positive samples confirmed by the BD Probe Tech diagnostic method were collected in Jersey and sent to Southampton. All 192 of those samples have been amplified using the CT1-CT5, VD2-CT5, CT1-VD3 or the P1-OMP2 primer combinations. Of those 192 samples, 40 are *ompA* PCR positive by the method outlined.

18 PCR products of the 40 *C.trachomatis ompA* PCR positive samples have been sequenced using the internal VD4 primer and external CT5 primer. The resulting sequences were compared with those published for variable domain four by Yuan *et al.,* (1989). These are listed in table 4.1.

Table 4.1: Sequencing results of VD4 genotyped samples received from Jersey during 2005

Genotype identification	Number of clinical samples
D	1
E	2
F	1
G	3
J	2
K	1
Unconfirmed	2

The two samples that have not been assigned to a genotypic grouping are both suspected to be of the D genotype however, the sequencing data was not of a quality to complete the serovar assignment with confidence. Both of these samples have been sequenced on several occasions however, the data has not

improved and regrettably it is thought that the template DNA in these two samples has degraded due to repeat freeze thawing of the samples.

Figure 4.2: A comparative alignment of variable domain 4 **sequence data from 2005 samples**

The Yellow highlights are characteristics typical of a serovar isolate.

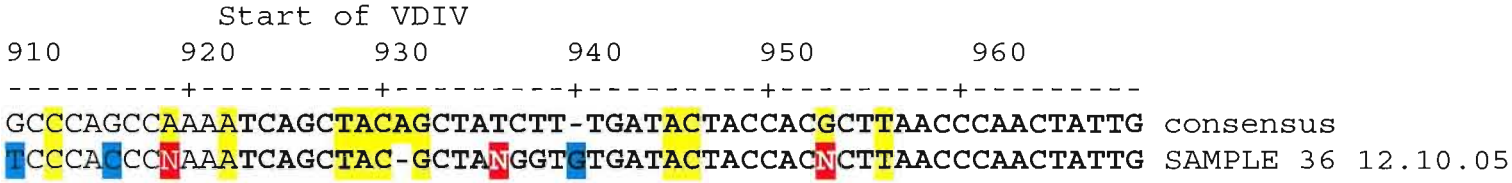
The Blue highlights are mis - matches within the aligned nucleotide sequences.

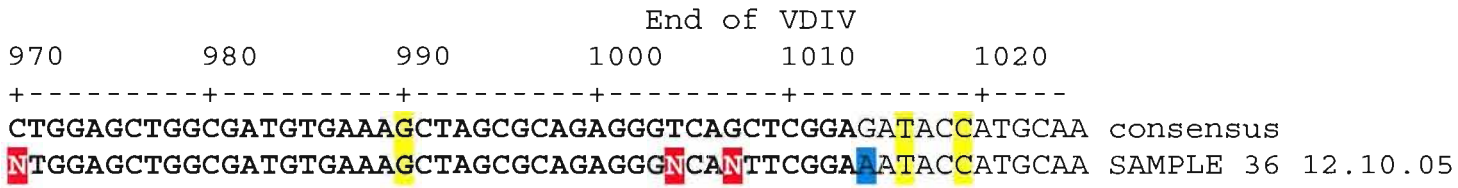
The Red highlights are errors within the aligned nucleotide sequences.

Serovar D

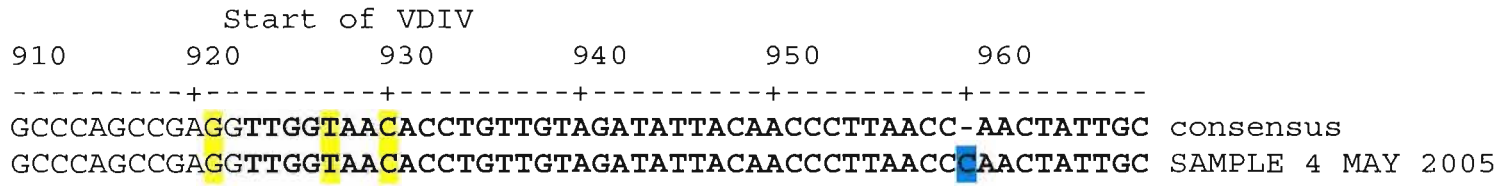
```
           Start of VDIV
910       920        930        940        950        960
---------+---------+---------+---------+---------+---------
GCCCAGCCAAAATCAGCTACAGCTATTTTT-GATAC-TACCACGCTTAACCCAACTAT consensus
TTCCACCCAAAATNAGCTANCGG-CNNNGNNGATAC-TACCATGCTTAACCCAACTAT SAMPLE 37 12.10.05
GCCCAGCCAAAATCAGCTACAGCTATTTTT-GATACNTACCACGCTTAACCCAACTAT SAMPLE 43 12.10.05 (UNCOMFIRMED)
          AAATNAGNTACNGGGANGGTNTGATAC-TACCACNCTTAACCCAACTAT SAMPLE 45 12.10.05 (UNCOMFIRMED)


                                      End of VDIV
970       980        990        1000       1010       1020
+---------+---------+---------+---------+---------+----
TGCTGGAGCTGGCGAT-GTGAAAACTGGC-GC-AGAGGGTCAGC-TCGGAGACACAATGCAA consensus
TGNTGGAGCTGGCNAT-GTGAAAACTGGCAGC-NGAGGGTCANC-TCGGAAACACAATGCAA SAMPLE 37 12.10.05
TGCTGGAGCTGGCGAT-GTGAAAACTGGC-GC-AGAGGGTCAGC-TCGGAGACNCCNTGCAA SAMPLE 43 12.10.05 (UNCOMFIRMED)
TGNTGGAGCTGGCGATTGTGAANACTGGCTGCNAGNGGGNCACCCTCGGAAACCCAATGCTT SAMPLE 45 12.10.05 (UNCOMFIRMED)
```

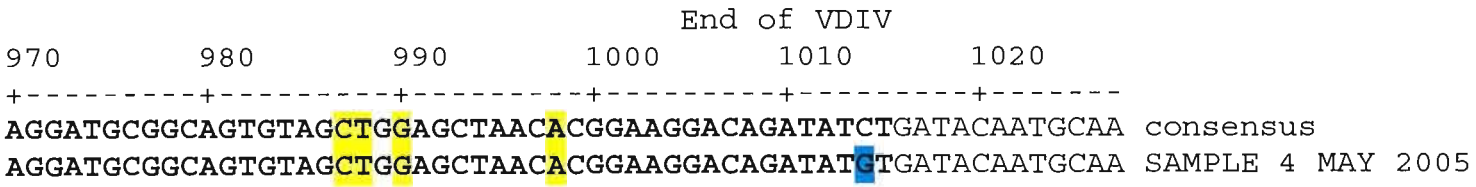
Serovar E



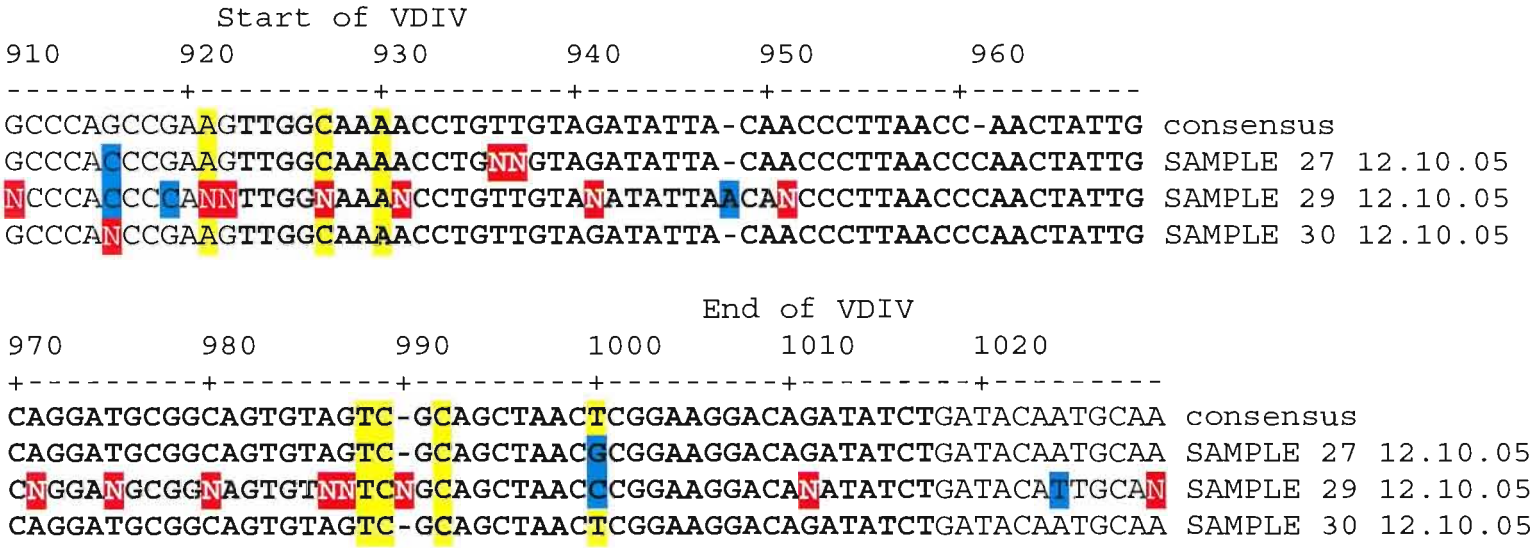


Serovar F





Serovar G


Start of VDIV
910 920 930 940 950 960
GCCCAGCCGAAGTTGGCAAAACCTGTTGTAGATATTA-CAACCCTTAACC-AACTATTG consensus
GCCCACCCGAAGTTGGCAAAACCTGNNGTAGATATTA-CAACCCTTAACCCAACTATTG SAMPLE 27 12.10.05
NCCCACCCCANNTTGGNAAANCCTGTTGTANATATTAACANCCCTTAACCCAACTATTG SAMPLE 29 12.10.05
GCCCANCCGAAGTTGGCAAAACCTGTTGTAGATATTA-CAACCCTTAACCCAACTATTG SAMPLE 30 12.10.05
End of VDIV
970 980 990 1000 1010 1020
CAGGATGCGGCAGTGTAGTC-GCAGCTAACTCGGAAGGACAGATATCTGATACAATGCAA consensus
CAGGATGCGGCAGTGTAGTC-GCAGCTAACGCGGAAGGACAGATATCTGATACAATGCAA SAMPLE 27 12.10.05
CNGGANGCGGNAGTGTNNTCNGCAGCTAACCCGGAAGGACANATATCTGATACATTGCAN SAMPLE 29 12.10.05
CAGGATGCGGCAGTGTAGTC-GCAGCTAACTCGGAAGGACAGATATCTGATACAATGCAA SAMPLE 30 12.10.05

Serovar J

```
            Start of VDIV
910       920        930        940        950         960
---------+---------+---------+---------+----------+--------
GCTCAGCCTAAATTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGC consensus
GCTCAACCTAAATTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGC SAMPLE 28 12.10.05
NTTCANCCTAAATTGGNTGANGCAANCTTGGATGTCACTACTCTAAACCCGACCATCGC SAMPLE 33 12.10.05

                                    End of VDIV
970        980        990        1000       1010       1020
+----------+---------+---------+---------+---------+------
TGGTAAAGGAACTGTGGTCG-CTTCCGGAAGCGAAAACGACCTGGCTGATACAATGCAA consensus
TGGTAAAGGAACTGTGGTCG-CTTCCGGAAGCGAAAACGACCTGGCTGATACAATGCAA SAMPLE 28 12.10.05
TGGTAAAGGAACTGTGGTCANCTTCCGGAAGCGAAAACGANCTGGCTGATTCAATGCAA SAMPLE 33 12.10.05
```

Serovar K

```
            Start of VDIV
910       920        930        940        950         960
---------+---------+---------+---------+----------+---------
GCTCAGCCTAAATTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCAC consensus
GCTCAACCTAAATTGGCTGAAGCAANNTTGGATGTCACTACTCTAAACCCGACCATCGC SAMPLE 6 12.10.05

                                    End of VDIV
970        980        990        1000       1010       1020
+----------+---------+---------+---------+---------+------
TGGTAAAGGAGCTGTGGTCTCTTCCGGAAGCGATAACGAACTGGCTGATACAATGCAA consensus
TGGTAAAGGAGCTGTGGTCTCTTCCGGAAGCGATAACGAACTGGCTGATACAATGCAA SAMPLE 6 12.10.05
```

Table 4.2: Sequencing results of VD4 genotyped samples received from Jersey during 2006

Genotype identification	Number of clinical samples
E	6
F	1
J	2
K	1
Unconfirmed	1

The unconfirmed sample noted during 2006 is suspected to be genotypically identical to the E grouping. However, this has not been possible to confirm completely due to the poor quality of the sequencing output.

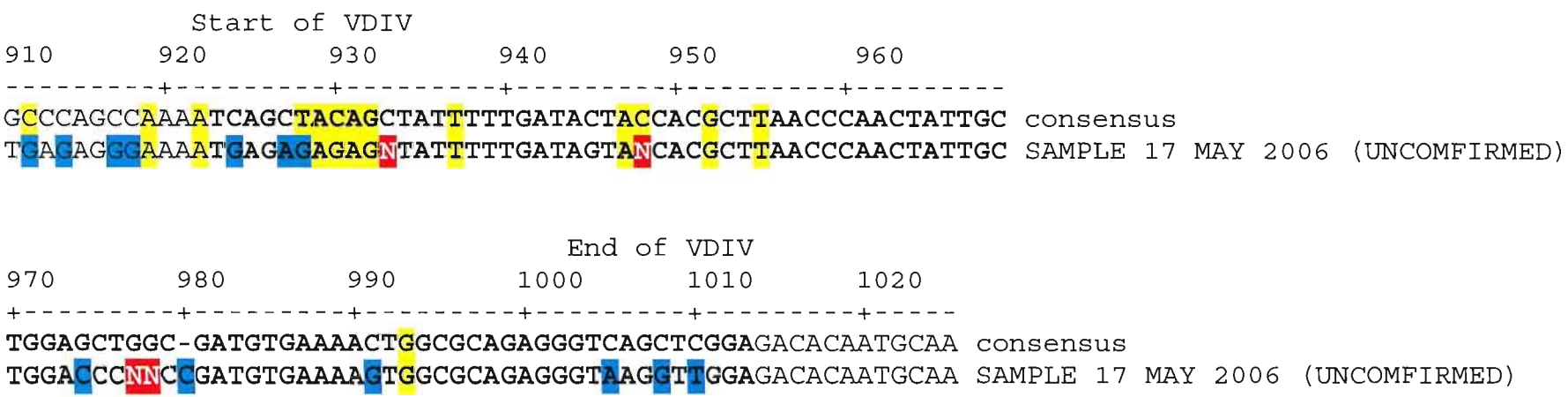
Figure 4.3: A comparative alignment of variable domain 4 **sequence data from 2006 samples**

The Yellow highlights are characteristics typical of a serovar isolate.

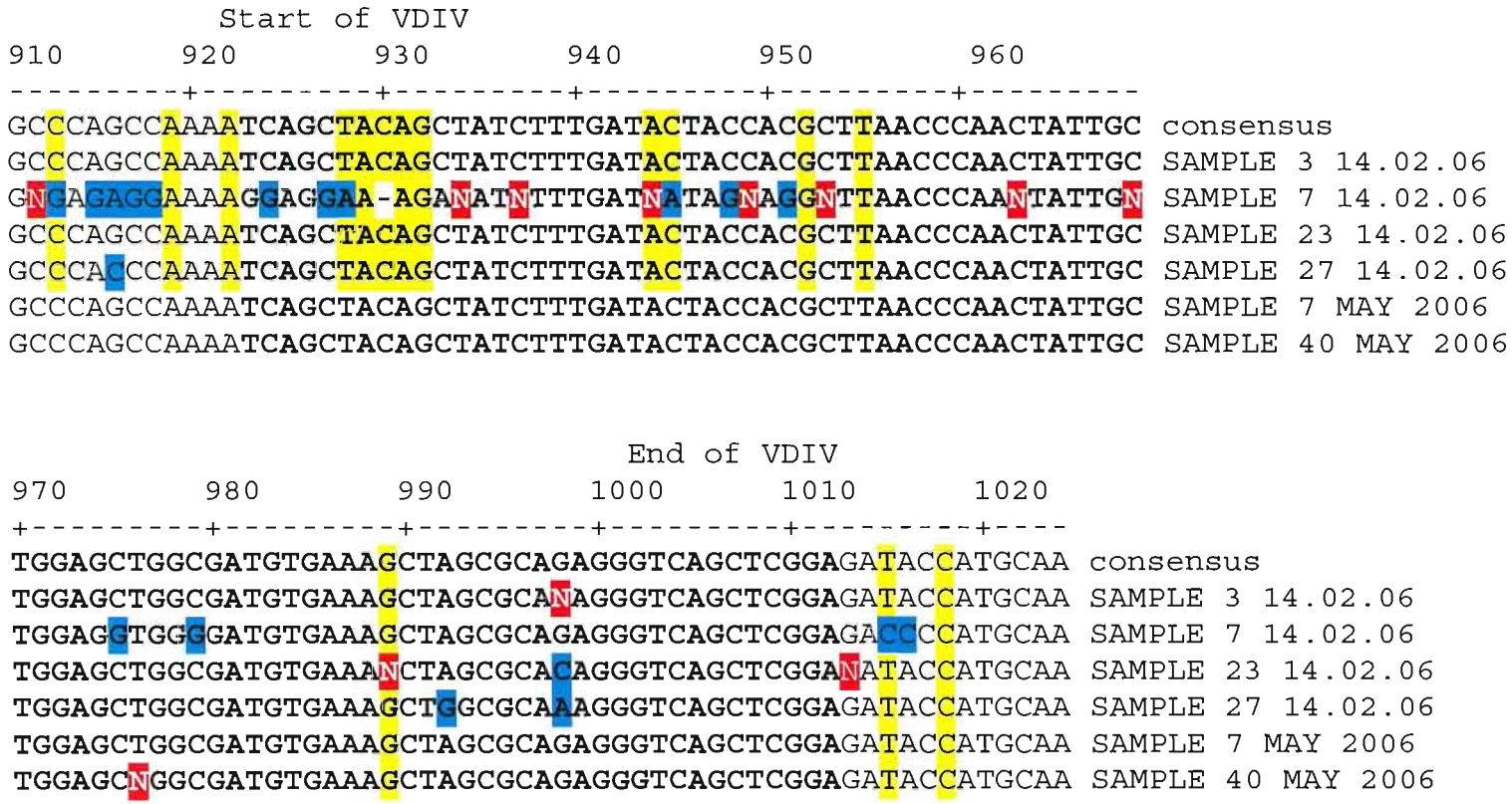
The Blue highlights are mis - matches within the aligned nucleotide sequences.

The Red highlights are errors within the aligned nucleotide sequences.

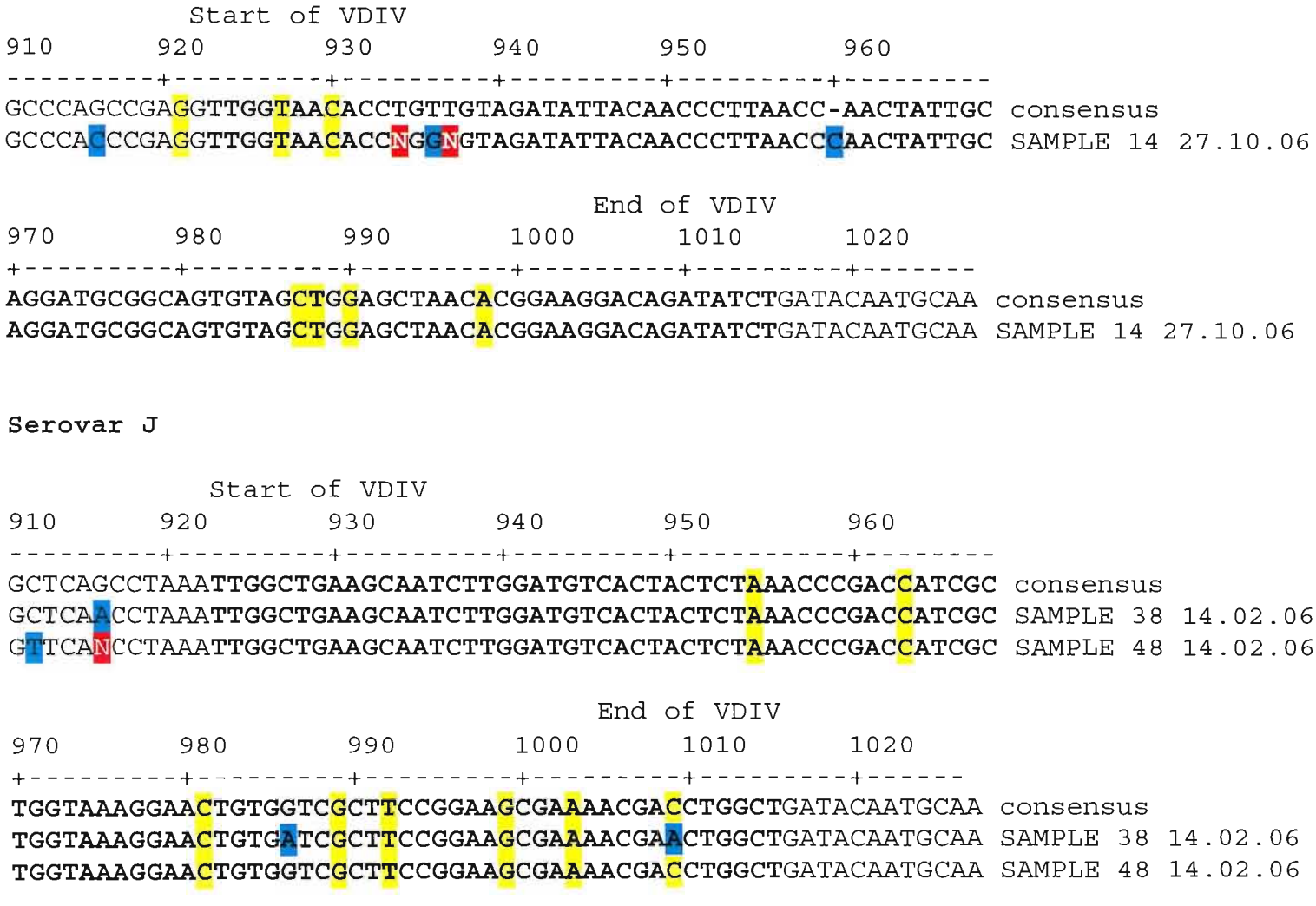
Serovar D



Serovar E


Start of VDIV
910 920 930 940 950 960
GCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC consensus
GCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC SAMPLE 3 14.02.06
GNGAGAGGAAAAGGAGGAA-AGANATNTTTGATNATAGNAGGNTTAACCCAANTATTGN SAMPLE 7 14.02.06
GCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC SAMPLE 23 14.02.06
GCCCACCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC SAMPLE 27 14.02.06
GCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC SAMPLE 7 MAY 2006
GCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGC SAMPLE 40 MAY 2006
End of VDIV
970 980 990 1000 1010 1020
TGGAGCTGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGATACCATGCAA consensus
TGGAGCTGGCGATGTGAAAGCTAGCGCANAGGGTCAGCTCGGAGATACCATGCAA SAMPLE 3 14.02.06
TGGAGGTGGGGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGACCCCATGCAA SAMPLE 7 14.02.06
TGGAGCTGGCGATGTGAAANCTAGCGCACAGGGTCAGCTCGGANATACCATGCAA SAMPLE 23 14.02.06
TGGAGCTGGCGATGTGAAAGCTGGCGCAAAGGGTCAGCTCGGAGATACCATGCAA SAMPLE 27 14.02.06
TGGAGCTGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGATACCATGCAA SAMPLE 7 MAY 2006
TGGAGCNGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGATACCATGCAA SAMPLE 40 MAY 2006

Serovar F



Serovar J

Serovar K

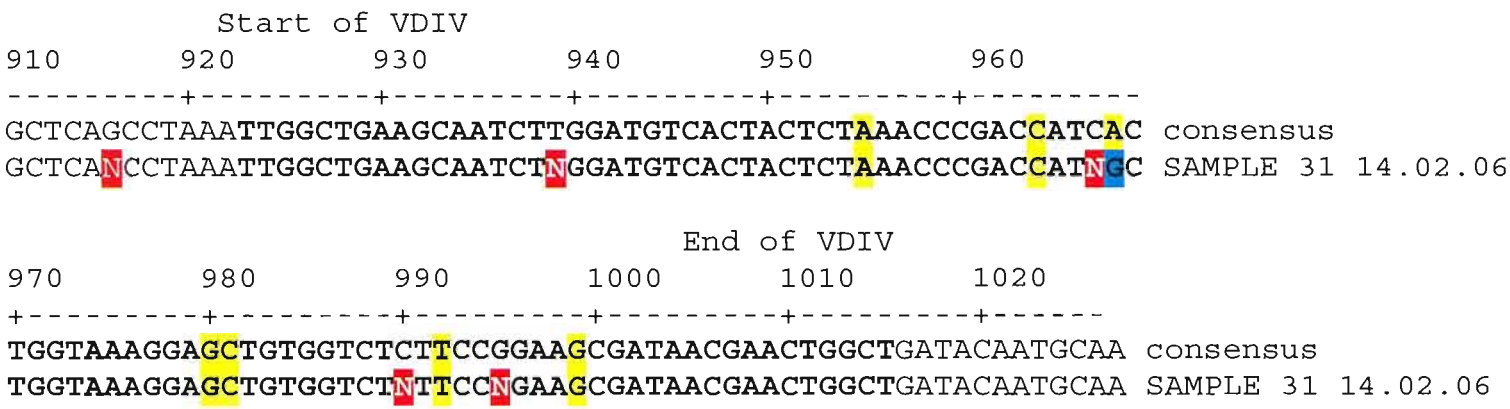

Start of VDIV
910 920 930 940 950 960
GCTCAGCCTAAATTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCAC consensus
GCTCANCCTAAATTGGCTGAAGCAATCTNGGATGTCACTACTCTAAACCCGACCATNGC SAMPLE 31 14.02.06
End of VDIV
970 980 990 1000 1010 1020
TGGTAAAGGAGCTGTGGTCTCTTCCGGAAGCGATAACGAACTGGCTGATACAATGCAA consensus
TGGTAAAGGAGCTGTGGTCTNTTCCNGAAGCGATAACGAACTGGCTGATACAATGCAA SAMPLE 31 14.02.06

Table 4.3: Sequencing results of VD4 genotyped samples received from Jersey during 2007

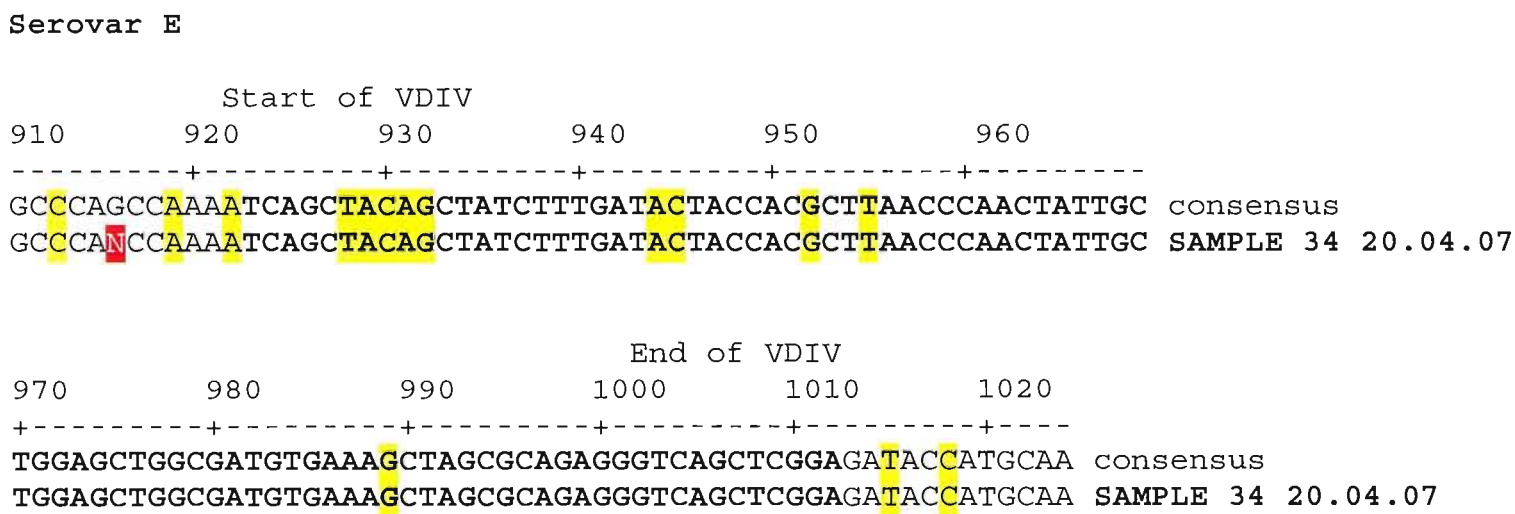
Genotype identification	Number of clinical samples
E	1
F	1

Figure 4.4: A comparative alignment of variable domain 4 sequence data from 2007 samples

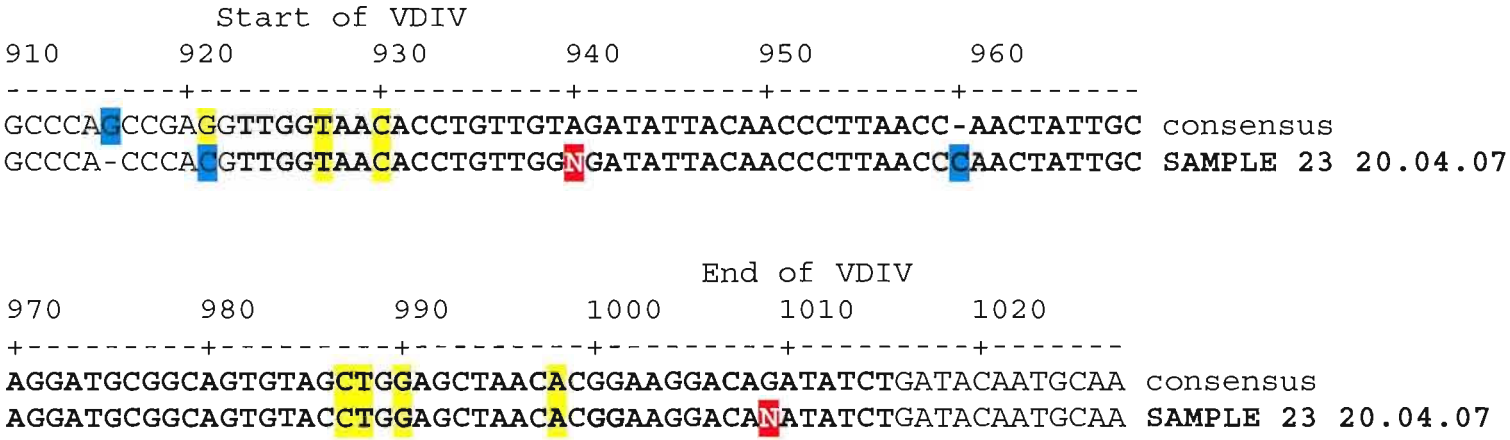
The Yellow highlights are characteristics typical of a serovar isolate.

The Blue highlights are mis - matches within the aligned nucleotide sequences.

The Red highlights are errors within the aligned nucleotide sequences.



Serovar F


Start of VDIV
910 920 930 940 950 960
GCCCAGCCGAGGTTGGTAACACCTGTTGTAGATATTACAACCCTTAACC-AACTATTGC consensus
GCCCA-CCCACGTTGGTAACACCTGTTGGNGATATTACAACCCTTAACCCAACTATTGC SAMPLE 23 20.04.07
End of VDIV
970 980 990 1000 1010 1020
AGGATGCGGCAGTGTAGCTGGAGCTAACACGGAAGGACAGATATCTGATACAATGCAA consensus
AGGATGCGGCAGTGTACCTGGAGCTAACACGGAAGGACANATATCTGATACAATGCAA SAMPLE 23 20.04.07

Because the *ompA* specific positive sample numbers were so low at 13.2%, it was not possible to draw conclusive evidence to represent a true reflection of the molecular epidemiology on Jersey during the 3-year study.

CHAPTER 5 COMPARATIVE ANALYSIS OF DIAGNOSTIC PLATFORMS

5.0 *OmpA* screening of Jersey and Southampton samples

Of the samples received from Jersey positive for *C.trachomatis* by the BD Probe Tech system, only approximately 10% were positive using the in-house *ompA* specific CT1-CT5 PCR method. To investigate whether this was typical of other published studies (where commercial NAATs systems had been used to screen samples and determine *C.trachomatis* positivity) the in-house PCR method was compared to commercial NAATs. A literature review was completed and a small project to test samples from the Southampton HPA laboratory was devised. Two different technologies for the detection of *C.trachomatis* were available, strand displacement amplification via the BD Probe Tech *C.trachomatis* detection platform (Jersey) and Taqman based PCR via the Roche Cobas Taqman based *C.trachomatis* detection system, (Southampton HPA). It was necessary to see if any other studies had used these systems as a diagnostic platform for the development of a PCR method focused on an *ompA* specific PCR product.

Ngandjio *et al,.* (2003) described the use of the Roche Cobas Amplicor system as a screening platform to determine *C.trachomatis* positivity in samples collected from a volunteer student population in Cameroon. Of the 1,277 students who voluntarily gave samples the prevalence of *C.trachomatis* infection was reported as 3.78%, as 48 of the original 1,277 students were deemed to have tested positive for *C.trachomatis.* Within this small group only 25 samples could be verified by further sequencing to determine the genotype from the *ompA* gene (i.e. approximately 50% of the known positives).

During the literature search, many studies were found highlighting the comparative merits of the BD Probe Tech and Roche Cobas Taqman based

C.trachomatis diagnostic systems. However, most of these studies did not use both methods and were direct comparative analyses of the platforms to other diagnostic methods but not to in-house PCR and sequencing for *ompA*. These methods did not involve the development or use of an independent PCR based method to then clarify the results from one or both of the diagnostic platforms. A comparison that assessed the performance of the BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic systems, was completed by Van Dyck *et al.,* (2001). This study collected 733 endocervical swab specimens from commercial sex workers in a multi-centred study. Samples were collected in Cotonou, Benin and Durban in South Africa and in Hat Yai, Thailand. All 733 samples were tested using the BD Probe Tech and Roche Cobas Amplicor based *C.trachomatis* diagnostic systems. Of those 733 samples tested 43 were found to be positive for *C.trachomatis* by both methods (5.8%). They found that although the diagnostic methods had different plasmid based targets, one was based on SDA technology and the other was based on PCR; both methods had similar sensitivity (94.0 and 98.0) and specificity (100 and 98.0) in the detection of *C.trachomatis*.

Another study (Chan *et al.,* 2000) also comparatively analysed the BD Probe Tech and Roche Cobas Amplicor based *C.trachomatis* diagnostic systems, however, in this study there was direct detection of *C.trachomatis* in male and female urine samples. 825 male and 399 female first-void urine specimens were collected during the study from three sexually transmitted infection clinics and doctors' offices located in Saskatchewan, Canada. The study found that the BD Probe Tech and Roche Cobas Amplicor based *C.trachomatis* diagnostic systems showed 163 and 162 positive *C.trachomatis* samples respectively (14%) from the 1,154 samples tested. The study found that the overall sensitivity of the BD Probe Tech and Roche Cobas Amplicor based *C.trachomatis* diagnostic systems, was 95.9% and 95.3% respectively. The specificity of the BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic systems, was calculated to be 99.3% and 98.3%. Chan *et al.,* (2000) concluded that the performance rates of the BD Probe Tech *C.trachomatis* diagnostic system, were similar to

those of the Roche Cobas Amplicor based *C.trachomatis* diagnostic system. Although the paper suggests the intention of an in-house plasmid based PCR for the evaluation of discrepancy results there was no data published by Chan *et al.,* (2000).

To date no study has been found that directly compares BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic systems with an *ompA* based in-house PCR and sequencing. There are publications that used plasmid based in-house PCR to double check the positivity of samples showing discrepancies however, none have gone on to either develop or use a published method that targets *ompA* (Chan *et al.,* 2000). It was decided therefore that a comparative analysis should be completed of the BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic systems with an *ompA* based in-house PCR.

One hundred first-void urine samples were collected in the Southampton laboratory after testing positive for *C.trachomatis* with the Roche Cobas Taqman based *C.trachomatis* diagnostic system. In addition the DNA extract required for Roche cobas Taqman based *C.trachomatis* diagnostic system was also collected. All 100 DNA extract samples were amplified using the CT1-CT5 primers. It was intended that any positive PCR amplicons would be genotyped via sequencing using the VD4 and CT5 primers. 14 samples were positive by PCR but unfortunately, none of the samples gave products of a quality that could be used for *ompA* sequence genotyping.

It was intended that 100 first-void urine samples would be collected in the laboratory in Jersey and screened via the BD Probe Tech based *C.trachomatis* diagnostic system and in addition through the Roche Cobas Taqman based *C.trachomatis* diagnostic system as a comparative analysis of the two platforms. However, only 27 first-void urine samples have been collected from Jersey and

processed. All of these samples proved positive using the BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic system and negative when amplified with the CT1-CT5 primers.

Because of the low numbers of samples collected in Jersey and due to constraints in time and as a reflection of the differences in population size between there and Southampton, it was decided to look at the samples already collected in previous years and in -80°C storage. Of the 341 clinical samples collected in Jersey, 45 proved positive by PCR however none of those positive samples were first-void urine samples. There were first void - urine samples collected during the previous 3 years however, whilst these samples were positive by the BD Probe Tech based *C.trachomatis* diagnostic system in Jersey, all were negative for *C.trachomatis* when amplified using the CT1-CT5 primers.

The *ompA* specific *C.trachomatis* positivity rate of 13.2% found in this study using the CT1-CT5 primers, was initially thought to be very low. It is essential to take into account that the work of Ngandjio *et al.,* (2003), they used samples that were swabs, taken for cell culture and the *ompA* amplification was applied to 'isolates' that had been amplified by culture. This is not a true comparison to the results here. This is because the samples collected and amplified by the in-house PCR, for this study were not amplified by culture and so represented quantities of *C.trachomatis* bacteria collected during routine diagnostic testing with the BD Probe Tech and the Roche Cobas Taqman based *C.trachomatis* diagnostic systems.

5.1 New *C.trachomatis* Swedish variant deletion in the cryptic plasmid

During the course of the screening of samples in Jersey via the BD Probe Tech *C.trachomatis* detection assay, a possible new strain of *C.trachomatis* was brought to our attention. This new variant *C.trachomatis* (nvCt) strain evaded detection by the Roche Cobas Taqman based *C.trachomatis* detection assay because of a 337bp deletion in the cryptic plasmid where the Roche diagnostic Taqman primers and probes were targeted (Ripa & Nilsson, 2006). In the Southampton Diagnostic laboratory *C.trachomatis* screening was completed by the Roche Cobas Taqman system and so possible samples containing nvCt strain might have gone undetected. In contrast, the diagnostic laboratory in Jersey used the BD Probe Tech *C.trachomatis* detection assay. This meant that we were in a position to screen for possible samples containing this nvCt strain because if it were present it would be detected by the BD Probe Tech *C.trachomatis* diagnostic assay but not the Roche assay. In addition, in anticipation of the strain arriving in the UK via sexual networking within Europe it was essential to see if the strain was already present in Jersey.

To prepare the control tests an isolate of nvCt was sourced from Sweden and grown in cell culture using $M^{c}Coy$ cells. A T25 flask was inoculated and cultured according to method 2.5-2.5.1, the monolayer was harvested as described in the methods sections 2.7-2.7.1. A dilution series was made ranging from 10^{-2} to 10^{-10} of the original inoculum. 200µL of each dilution was inoculated into 200mL of sterile filtered urine. The urine had already been screened via the Roche Cobas Taqman based diagnostic assay and via the CT1-CT5 *ompA* specific PCR to clarify that it was negative for *C.trachomatis*. As a positive control the same method was completed using a L2-434/Bu which is known not to have a deletion in the cryptic plasmid from the sequencing data published by Commanducci *et al.,* (1998). Both dilution series were analysed through the BD Probe Tech and Roche Cobas Taqman based *C.trachomatis* diagnostic assays. As predicted both the L2-434/Bu and the nvCt were detected over the complete

range of dilutions using the BD Probe Tech *C.trachomatis* diagnostic assays. However, the Roche Cobas Taqman based *C.trachomatis* diagnostic assay only detected the L2-434/Bu. None of the urine samples seeded with the nvCt resulted in the detection of *C.trachomatis* by the Roche Cobas Taqman diagnostic platform. This meant that the results of Ripa & Nilsson, (2006) were confirmed, the nvCt strain could not be detected by the Roche Cobas Taqman based *C.trachomatis* diagnostic assay. It was also confirmed that output from the BD Probe Tech *C.trachomatis* diagnostic assay could be used up to a week later on samples analysed through the Roche Cobas Taqman based *C.trachomatis* diagnostic assay as positive or negative for *C.trachomatis*. Thus the screening protocol was validated.

It was hoped to collect a representative set of 100 clinical urine samples in Jersey. However, only 27 urine samples were collected from Jersey, due to constraints in time. These samples were divided in Jersey, at the time of collection into two separate samples. One of the samples was processed via the BD Probe Tech system to determine if it was *C.trachomatis* positive. Those samples that were *C.trachomatis* positive would have the other corresponding half of the sample couriered to the Southampton laboratory for *C.trachomatis* testing using the Roche Cobas Taqman platform. Those samples that were not positive for *C.trachomatis* via the BD Probe Tech method were discarded.

Samples received from Jersey were processed within three days of arrival at the Southampton laboratory (within the one week limit established by the original dilution series of L2-434/Bu). However, most were processed within 24 hours of arrival so as to be as close as possible to the timescales indicated by the Roche Cobas Taqman based *C.trachomatis* diagnostic assay.

All 27 collected BD Probe Tech *C.trachomatis* positive urine samples were analysed using the Roche Cobas Taqman platform. Of those 27 samples 25 were positive for *C.trachomatis*. This meant that in 25 of the samples there was no deletion in the cryptic plasmid i.e no nvCt strain.

The 2 samples that proved negative were retested using the Roche Cobas Taqman *C.trachomatis* diagnostic platform; this resulted in one of the samples giving a positive result the other remained negative. It was realised that it was necessary to test the samples with another method to ensure a reliable conclusion as to whether these samples contained the nvCt. A PCR based protocol, where primers had been designed that bridged the area of the 377bp deletion within the cryptic plasmid was employed (see section 2.10.1). The results of this showed that the two Roche Cobas Taqman *C.trachomatis* negative samples did not have the 377bp deletion in the cryptic plasmid, and therefore they were not from infections caused by the nvCt strain.

In addition to the 27 samples collected in Jersey, 100 samples were collected in Southampton. The samples were collected and screened using both the Roche Cobas Taqman and BD Probe Tech diagnostic platforms. All of the samples proved negative for nvCt infection.

CHAPTER 6 NEW VARIANT *C.TRACHOMATIS* FROM SWEDEN

6.0 *OmpA* genotyping of the nvCt

New variant *C.trachomatis* with a deletion in the cryptic plasmid was detected in Sweden following an apparent 25% decrease in *C.trachomatis* infections between November 2005 and August 2006 in Halland county, southwest Sweden (Ripa and Nilsson 2006). A sample of *C.trachomatis* with a suspected deletion in the cryptic plasmid was obtained from Malmo, Sweden and cultivated as described in the methods section 2.5 and 2.5.2 to provide a suitable inoculum. A genomic DNA extraction was completed (section 2.6-2.6.1) and the *ompA* PCR performed using the methods described by Ngandjio *et al.,* (2003). The CT1-CT5 (Ngandjio *et al.,* 2003) amplicons were sequenced in both directions using the VD4 and CT5 primers (Ngandjio *et al.,* 2003) and comparatively genotyped to a published sequence (Yuan *et al.,* 1989). The strain was confirmed to be genotypically identical to Serovar E, these results are highlighted in figure 6.1.

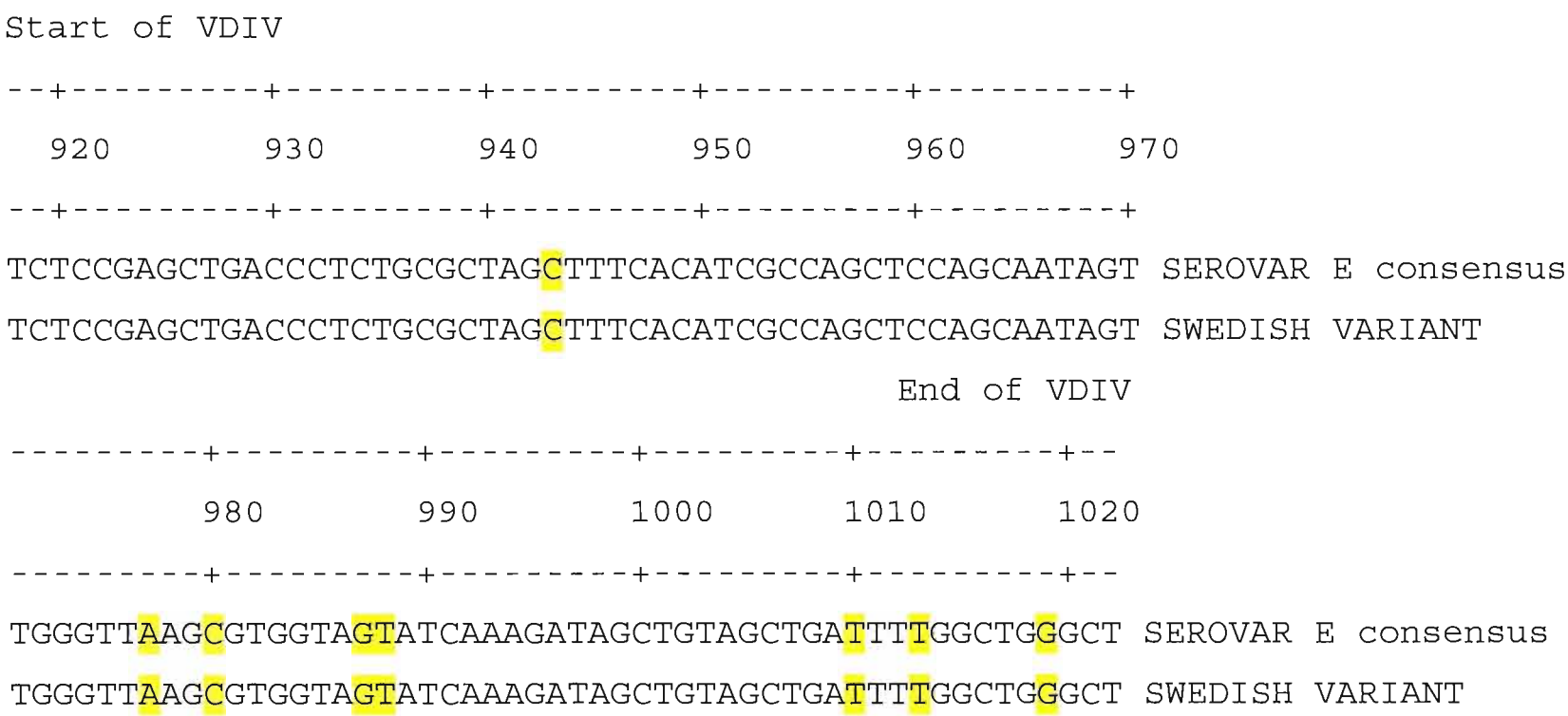


Figure 6.1: Sequence alignment showing the genotype E determinants of *ompA* variable domain IV from a published serovar E and the Swedish variant

The Yellow highlights are characteristics typical of a serovar E isolate.

6.1 Confirmation of the plasmid deletion

A part of the cryptic plasmid from the variant strain had already been sequenced. There is a deletion of 377 bps in the target area for the Abbott and Roche *C.trachomatis* NAAT tests. This deletion does not affect BD Probe Tech *C.trachomatis* test (Ripa & Nilsson, 2006). Figure 6.2 gives a representation of the location of the deletion

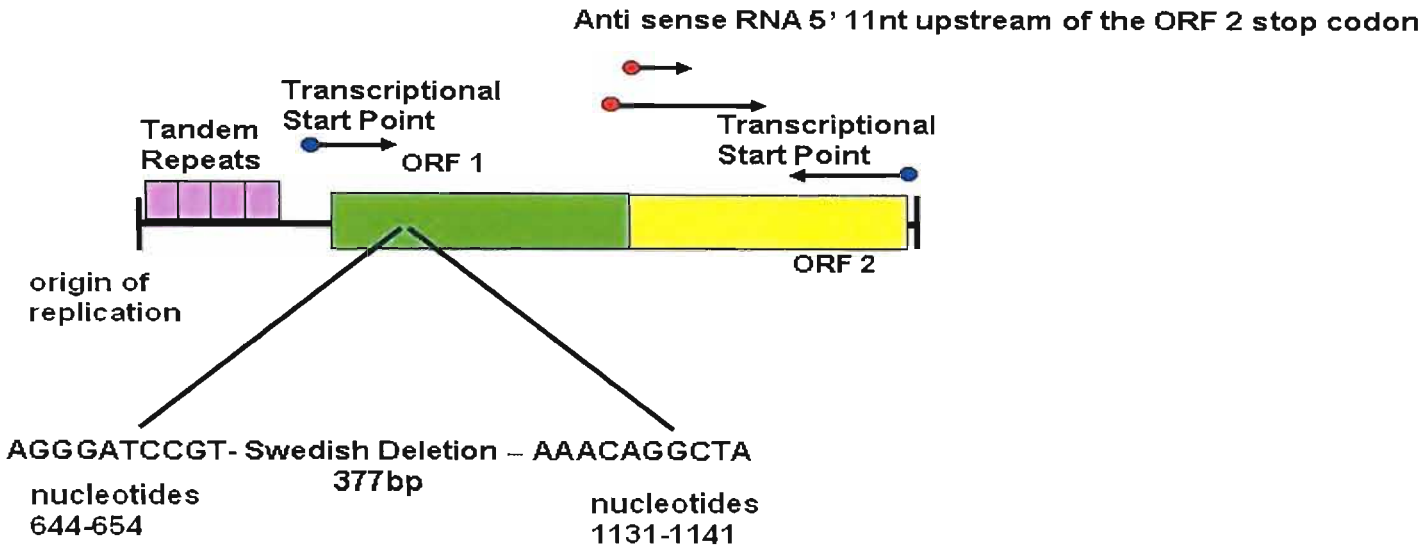


Figure 6.2: Location of the new variant 377bp deletion
A linear map of the *C.trachomatis* plasmid starting at the origin of replication and ending at the start codon of ORF 2. In addition a representation of the location of the 377bp deletion in the nvCt strain.

It was essential to confirm whether the strain sourced from Malmo had the suspected 377bp deletion in the cryptic plasmid therefore primers were used in a PCR reaction that mapped either side of the suspected deletion (Ripa & Nilsson, 2006). Optimisation of the amplification reaction was essential to ensure the highest quality data. This included variations in $MgCl_2$ concentrations.

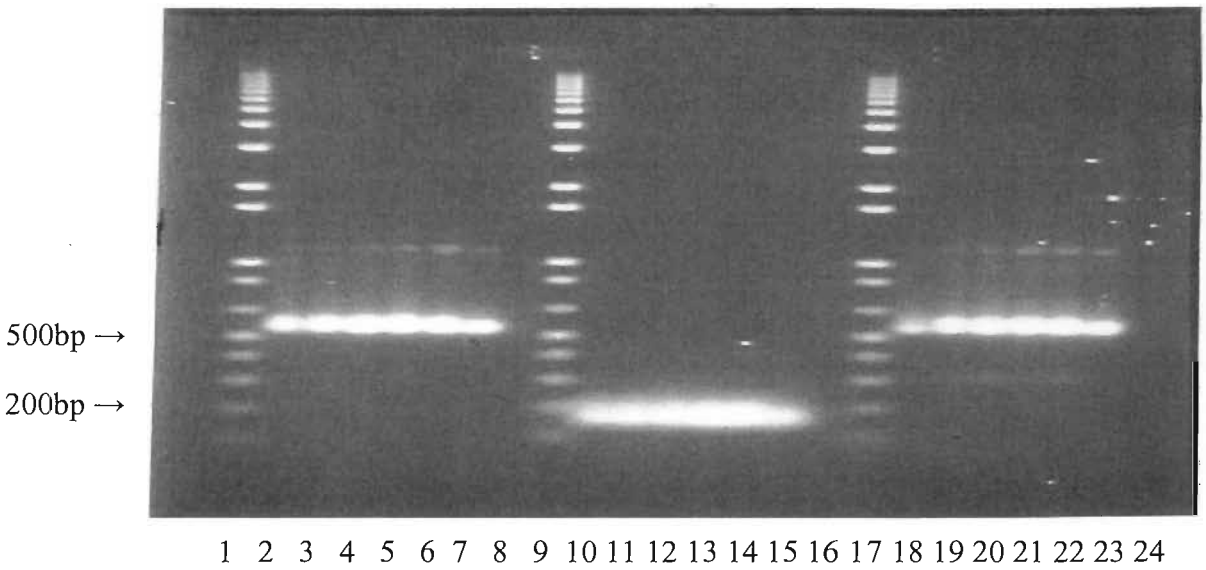


Figure 6.3: Optimisation of $MgCl_2$ for the detection of the swedish variant 377bp deletion in the cryptic plasmid.

Lanes 1, 9 and 17: 1 kb plus ladder, Lanes 2, 10 and 18:1.25mM $MgCl_2$ Lanes 3,11, and 19: 2.5mM $MgCl_2$, Lanes 4, 12, and 20: 3.75mM $MgCl_2$, Lanes 5, 13, and 21: 5mM $MgCl_2$, Lanes 6, 14, and 22: 6.25mM $MgCl_2$, Lanes 7, 15, and 23: 7.5mM $MgCl_2$, Lanes 8, 16, and 24:0mM, $MgCl_2$, Lanes 2-7 L1 DNA template, Lanes 10-15 New variant Swedish deletion isolate DNA template, Lanes 18-23 L3 DNA template,

The L1 and L3 samples each have bands (of 593bp) above the 500bp marker on the 1 kb plus ladder shown by lanes 1 to 7 and lanes 18 to 23, this is the expected size of the amplicon when the template DNA in the PCR reaction does not present with the 377bp deletion. However, lanes 10 to 15 show a band (of 216bp) approximately at the 200bp marker on the 1 kb plus ladder this would

correspond to the 377bp that is expected to be missing from this amplicon if the template DNA contained this specific deletion in the cryptic plasmid thus confirming that the Swedish strain does have the 377bp deletion in the cryptic plasmid and this was clearly shown on the gel image.

Urine samples that have been diagnostically tested for *C.trachomatis,* and confirmed positive using the BD Probe Tech assay have been transported to the Southampton diagnostic laboratory to be further tested by the Roche Cobas Taqman diagnostic assay for the new Swedish variant strain with the deletion in the cryptic plasmid. To date all the samples have proven positive by both methods of diagnostic detection and so it is concluded that the new Swedish variant has not reached Jersey yet (see chapter 5). Screening will continue however in anticipation of this happening in the future.

CHAPTER 7 MOLECULAR GENOTYPING OF PLASMIDS

7.0 Comparative analysis of the plasmids

The study of the new Swedish variant strain prompted investigation into finding the parental strain. Contact was made with the laboratory in Malmo, Sweden, to locate a possible parental isolate that could be cultured in Southampton and then the DNA extracted and the plasmid DNA amplified so that it could be sequenced and comparatively analysed to other published plasmid sequences.

7.1 Plasmid sequence selection

To complete a comprehensive study of the plasmid from the new Swedish variant strain it was necessary to gather all of the *C.trachomatis* published plasmid sequences. Thirteen sequences were collected. However, it was known that a further two sequences had been completed (Stephens *et al.,* (1998): Black *et al.,* 1989)). The Authors of these sequences where contacted directly to try to obtain the plasmid sequences in an electronic format, as this would avoid unnecessary computing errors that could have arisen if the sequences were converted manually from a paper copy to an electronic format, but unfortunately neither sequence was obtainable in an electronic format and so both sequences were omitted from the analysis.

The 13 plasmid sequences that were gathered together were aligned to start at the origin of replication, as determined by Hatt *et al.,* (1988). This sequence was chosen as the comparative reference sequence as it was originally sequenced in Southampton and could be verified to the original sequencing data. In some cases it was necessary to reverse Complement and cut the sequences to complete the alignment from the origin of replication, this was completed using the Lasergene 6 package. Confirmation of the plasmid

alignment was completed using Bio edit and the Artemis and Artemis Comparison Tools (ACT) programmes.

Table 7.1 Plasmid sequence details for the alignment

Sequence	Accession number	Number of bps	Reverse Complemented	Cut at postion
A Carlson *et al.,* (2005)	CP000052	7510	YES	6397
CTA		7480	NO	3846
CTB		7502	NO	1879
B Sriprakash and Macavoy (1987)	M19487	7496	NO	6852
JALI		7506	NO	6141
D Commanducci *et al.,* (1990)	NC_001372	7524	NO	6858
E SW2		7169	YES	1212
E SW3		7502	YES	1212
L1 Hatt *et al.,* (1988)	X06707	7500	NO	
L2 Thomson *et al.,* (2008)		7499	NO	42
L2 Commanducci *et al.,* (1988)	X07547	7499	NO	6456
L2b London		7500	YES	4329
L2b Amsterdam		7528	YES	943

7.2 Quality of the aligned sequences and further selection

On completion of the alignment, it was noticed in the preliminary investigation that some of the sequencing data collected was not of a suitable quality to be used in the analysis as it was thought to be ridden with obvious sequencing and assembly mistakes. This data was also omitted from the comparison as it made the alignment misleading. The sequences that were omitted due to poor quality

were the serovar B sequence (Sriprakash & Macavoy 1987) and the L2b Amsterdam sequence, sequenced by the Sanger Centre. On contacting the Sanger Centre it was noted that the L2b Amsterdam sequence was not complete as the assembly of the sequence was only preliminary further checks were required.

On closer inspection it was noticed that the two published L2 sequences, one by Thomson *et al.,* (2008) accessed from the Sanger centre website and one by Commanducci *et al.,* (1988), were identical. It was thought to be unnecessary therefore, to include both sequences in the alignment as they showed identical data. The Commanducci *et al.,* (1988) was chosen to represent the L2 sequence as it was the most accessible due to an allocated accession number on publication.

7.3 Intergenic analysis from the origin of replications

The final 10 selected plasmid sequences were aligned from the origin of replication; this is usually identified by four 22bp tandem repeats in the sequence. It was noted however, that the CTA sequence sourced from the Sanger Centre was missing one of these repeats in its sequence and thus only has three of the usual four 22bp tandem repeats. This was not a feature of the published serovar A sequence (Carlson *et al.,* 2005). This missing sequence was intergenic and so does not affect the sequence of amino acids or any of the resulting proteins made during translation.

The published serovar A sequence (Carlson *et al.*, 2005) did however have a 8bp insertion starting at position 80 in relation to the L1 sequence (Hatt *et al.,* 1988). Again this was not a feature of the serovar A sequence source from the Sanger Centre and was thought to be questionable as a possible assembly error

as none of the other sequences showed any identity with this 8bp insertion and it is very similar to the sequence upstream of the insertion.

At position 111 there is a single nucleotide substitution from adenine in those serovars usually associated with ocular and genital tract infections to guanine in the LGV sequences. In addition at position 120 there is also a nucleotide substitution from cytosine to adenine in the serovar D sequence published by Commanducci *et al.,* (1990). Both of these bp substitutions are intergenic and so do not affect any resulting proteins predicted to be made during translation.

7.4 ORF 1

Overall 17 different changes in the 10 aligned sequences were noted in ORF 1 these include substitutions, deletions, and insertions some of which cause changes in the translated sequence and the length of the resulting protein.

The table below shows those changes to the sequences that do no result in a change to the translated sequences

Table 7.2 Changes in ORF 1 that do not result in a change to the translated sequence (non-synonymous mutations)

Sequences with the bp change	Type of change	Position	bp change
CTA, CTB and JALI	Substitution	194	A - C
D, SW2 and SW3	Substitution	200	C - T
D, SW2 and SW3	Substitution	398	C - T
D, SW2 and SW3	Substitution	410	C - A
L1, L2 and L2 LON	Substitution	785	A - G
L1, L2 and L2 LON	Substitution	923	C - T
A, CTA,CTB and JALI	Substitution	1004	T - C

At position 246 a single nucleotide substitution is noted from cytosine to thymine in the sequences of those serovars usually associated with genital tract infections, there is an amino acid change from histidine which is a charged hydrophilic amino acid to tyrosine a hydrophobic amino acid. This may result in a change to the folding of the resulting protein.

At position 646 a 22bp insertion is noted in the serovar D sequence (Commanducci *et al.,* 1990) this would obviously result in a change to the translation of the sequence to amino acids and results in a shortened protein being made during translation.

From position 653 to 1030 the serovar E sequence, SW2, shows a 377bp deletion in the coding sequence. This is the plasmid sequence of the nvCT isolate sourced from Malmo, Sweden. The 377bp deletion results in a considerably shorter ORF 1 protein.

At position 805 the JALI serovar B sequence shows a single nucleotide substitution from guanine to adenine which does result in a change to the amino acid sequence from cysteine to tyrosine which has a much larger R group because of the aromatic ring and so the folding of the protein may be affected.

Position 894 shows a single bp change in the coding sequence between the serovars usually associated with oculo-gential tract infections and those associated with LGV. The point mutation is a substitution from cytosine to thymine and results in an amino acid change from histidine which is hydrophilic to tyrosine known to be hydrophobic, thus there may be a change in the folding of the protein produced to accommodate this point mutation and resulting effects.

At position 910 in the L2 sequences there is a 1bp adenine deletion, it is thought that this bp deletion would be responsible for a premature termination in the resulting protein at position 948 rather than at position 1082.

Position 1051 shows a single nucleotide substitution from guanine to thymine in those strains usually associated with genital tract infections this results in an amino acid change from serine to isoleucine but it is unlikely that the folding of the protein will be affected due to a similarity in properties and size of these amino acids. The substitution would have had no effect on L2 at this point because of the premature termination of the ORF caused by the adenine deletion at position 910.

There is a single nucleotide substitution a position 1071 from adenine to cytosine in those sequences usually associated with LGV infection. It results in an amino acid substitution from leucine to proline but only in the L1 protein sequence this is because of the premature termination in the L2 ORF 1 sequences caused by the deletion of a single adenine at position 910. The last base and end of ORF 1 is located at position 1082 for all sequences except the L2 serovars.

There is one intergenic single nucleotide substitution between the end of ORF 1 and the end of ORF 2 at position 1126. It is a change from cytosine to thymine in those sequences usually associated with LGV infection but because of its intergenic nature no amino acid change was noted.

7.5 ORF 2

In total 11 changes were noted between the 10 different sequences of the alignment in ORF 2. To not confuse the alignment of the sequences the non coding strand of ORF 2 was used to report the changes in the sequence. In addition the numbering of nucleotides from the L1 sequence was continued; this meant that although the numbering was incremental the ORF was analysed from the stop codon backwards. Had there been any deletions in the sequences this would have been a problem however, all of the changes noted in ORF 2 are single nucleotide substitutions. The end of ORF 2 is located at position 1132 in all of the 10 aligned sequences.

Table 7.3 shows those changes to the sequences that do no result in a change to the translated sequences

Table 7.3 Changes in ORF 2 that do not result in a change to the translated sequence (non-synonymous mutations)

Sequences with the bp change	Type of change	Position	Bp change
A	Substitution	1160	G - A
A, CTA,CTB and JALI	Substitution	1217	T - C
D, SW2 and SW3	Substitution	1301	C - T
A, CTA,CTB and JALI	Substitution	1607	A - G
L1, L2 and L2 LON	Substitution	1637	T - C
L1, L2 and L2 LON	Substitution	1664	C - T
L1, L2 and L2 LON	Substitution	1856	G - A
A, CTA,CTB and JALI	Substitution	1904	G - A
D, SW2 and SW3	Substitution	2075	G - A

At position 1147 the bp substitution is from thymine to guanine in those genotypes usually associated with LGV, SW2 and SW3. This results in an amino

acid change from leucine to methionine. Both these amino acids have similar side chain properties and so it is unlikely to cause a change to the folding of the resulting protein.

There is one intergenic single nucleotide substitution between the start of ORF 2 and the end of ORF 3 at position 2153. It is a change from thymine to cytosine in those sequences usually associated with ocular infection but because of its intergenic nature no amino acid change was noted.

At position 2225 there is a 44bp insertion that is unique to the SW2 sequence it is an identical repeat. Initially, it was thought to be an error in the assembly of the sequence but the data was re-confirmed by the Sanger centre and is thought to be a true reflection of the sequence.

7.6 ORF 3

Overall 14 differences in the 10 aligned sequences were noted in ORF 3: these were all substitutions. Table 7.4 shows those changes to the sequences that do no result in a change to the translated sequences.

Table 7.4 Changes in ORF 3 that do not result in a change to the translated sequence (non-synonymous mutations)

Sequences with the bp change	Type of change	Position	Bp change
A, CTA, CTB and JALI	Substitution	2355	T – C
A, CTA, CTB and JALI	Substitution	2583	T – C
L1, L2 and L2 LON	Substitution	2676	T – C
JALI	Substitution	2832	A – G
D, SW2 and SW3	Substitution	2901	G – A
L1, L2 and L2 LON	Substitution	2919	A – T
D, SW2 and SW3	Substitution	3114	A – G
D, SW2 and SW3	Substitution	3165	T – C
L1	Substitution	3339	T – C

In the serovar D sequence published by Commanducci *et al.,* (1990) at position 2386 there is a single nucleotide substitution from a thymine to a cytosine that subsequently causes a tyrosine to histidine change in the resulting amino acid sequence. Because both these amino acids have similar properties in that they are both polar and have aromatic rings it is unlikely that there will be a significant change in the tertiary structure of the resulting protein.

In serovars usually associatied with genital tract infection at position 2786 in ORF3 there is a nucleotide substitution from cytosine to thymine that results in an amino acid change from alanine to valine. Both alanine and valine are non

polar, aliphatic amino acids and so because of their carbon - hydrogen based side chains are considered non-reactive. In addition, because of the hydrophobicity of valine it is likely that it will be centralised in a hydrophobic core rather than be surface exposed in the resulting tertiary structure of the protein.

At position 2788 there is a single nucleotide substitution of a guanine to an adenine that results in an amino acid change from alanine to threonine in those genotypes usually associated with LGV infection. Both amino acids share the aliphatic and hydrophobic properties. Threonine is a small slightly polar amino acid and alanine is seen as a tiny slightly non polar amino acid, however both amino acids share the aliphatic and hydrophobic properties.

In those genotypes usually associated with LGV infection at position 2822, guanine is substituted for cytosine and this causes a single amino acid change from glycine to alanine in the protein. These amino acids are considered to have similar properties, both are classified as tiny, aliphatic, non polar, hydrophobic amino acids and because of their similar properties are unlikely to change the resulting tertiary structure of the protein.

At position 3185 in those genotypes usually associated with ocular infection there is a single nucleotide substitution from guanine to adenine that causes a single amino acid substitution from arginine to glutamine. Both amino acids are polar. Being polar, glutamine is generally surface exposed in the folding of a protein, and thus exposed to an aqueous environment. This also means it may have a significant role to play in the function of the protein.

7.7 ORF 4

Overall 11 different changes in the 10 aligned sequences were noted in ORF 4 these were all substitutions. Table 7.5 shows those changes to the sequences that do no result in a change to the translated sequences

Table 7.5 Changes in ORF 4 that do not result in a change to the translated sequence (non-synonymous mutations)

Sequences with the bp change	Type of change	Position	Bp change
L1, L2 and L2 LON	Substitution	3758	A - G
L1, L2 and L2 LON	Substitution	3764	C - T
L1, L2 and L2 LON	Substitution	4253	T - C
L1, L2 and L2 LON	Substitution	4418	C - T
L1, L2 and L2 LON	Substitution	4487	C - T
A, CTA, CTB, and JALI	Substitution	4508	C - T
L1, L2 and L2 LON	Substitution	4571	C - T
L1	Substitution	4586	T - C

At position 3798 in those genotypes usually associated with ocular infection there is a single nucleotide substitution from guanine to adenine that results in a single amino acid substitution from valine to isoleucine. As valine is an aliphatic, hydrophobic, amino acid it favours substitution with amino acids of similar properties such as isoleucine. Both valine and isoleucine have C-beta branching, meaning there is a lot more bulkiness near to the protein backbone, and thus that the resulting amino acid chain is more restricted in the conformation it can adopt. But because of the similar properties of these amino acids it is unlikely that there will be a significant change in the tertiary structure of the protein.

At position 4140 in those genotypes usually associated with genital tract infection there is a single nucleotide substitution from a thymine to a cytosine that results in a change to the amino acid sequence from a serine to a proline.

In those genotypes usually associated with LGV infection at position 4364 there is a single nucleotide substitution from adenine to cytosine that results in a leucine to phenylalanine change to the amino acid sequence. Both amino acids are hydrophobic and are therefore buried in protein hydrophobic cores. The leucine side chain like that of phenylalanine is very non-reactive, and is thus rarely directly involved in protein function though it can play a role in substrate recognition, it is possible that this maybe the function in this case.

In the SW2 genetic sequence there is a point mutation at position 4479. A substitution of cytosine to adenine results in an amino acid change from glutamine to lysine. Both amino acids are quite frequently involved in protein active or binding sites due to their polarity and their preference for exposure on the surface of a protein. This may possibly the role in this case.

There is one intergenic point mutation at position 4644. This is a substitution from guanine to adenine in those genotypes usually associated with LGV infection. Because of the intergenic nature of the substitution there was no resulting amino acid change.

7.8 ORF 5

Overall 14 different changes in the 10 aligned sequences were noted in ORF 5 these were all substitutions. Table 7.6 shows those changes to the sequences that do no result in a change to the translated sequences

Table 7.6 Changes in ORF 5 that do not result in a change to the translated sequence (non-synonymous mutations)

Sequences with the bp change	Type of change	Position	Bp change
A, CTA, CTB, and JALI	Substitution	5352	G - A
D, SW2 and SW3	Substitution	5466	G - A
A, CTA, CTB, and JALI	Substitution	5469	G - A

In those genotypes usually associated with LGV infection at position 4732 there was a single nucleotide substitution from guanine to cytosine that resulted in an amino acid change from glutamate to glutamine. Both amino acids very similar the only difference is that glutamine contains an oxygen in place of the amino group. Both are polar amino acids and thus prefer to be exposed on the surface of proteins where they frequently take an active role in binding sites. This may be the role in this case.

At position 4879 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from a thymine to a cytosine that results in a change to the amino acid sequence from a serine to a proline. Due to its small size serine is relatively common within tight turns on the protein surface, where it is possible for the serine side-chain hydroxyl oxygen to form a hydrogen bond with the protein backbone, effectively mimicking proline. Proline can often be found in very tight turns in protein structures (i.e. where the polypeptide chain must change direction), because of this, despite being

aliphatic and hydrophobic, proline is usually found on the protein surface. This change in direction of the protein may be the function of this amino acid in the protein structure.

In those genotypes usually associated with LGV infection at position 4967 there was a single nucleotide substitution from adenine to guanine that resulted in an amino acid change from aspartate to glycine. Aspartate is a negatively charged, polar amino acid and generally prefers to be exposed on the surface of proteins. When buried within the protein, aspartate is frequently involved in salt-bridges, pairing with a positively charged amino acid to create stabilising hydrogen bonds, this can be important for protein stability. Glycine contains a single hydrogen as its side chain (rather than a carbon as is the case in all other amino acids). This means that there is much more conformational flexibility in glycine and it can reside in parts of protein structures that are forbidden to all other amino acid. It is likely therefore that this substitution may have some impact on the tertiary structure of the protein.

At position 4994 in those genotypes usually associated with ocular infection there is a single nucleotide substitution from guanine to adenine that results in a change to the amino acid sequence from a serine to asparagine. Due to the similar properties of serine and asparagine there is likely to be little change to the tertiary structure of the protein as both are polar, non – charged amino acids. The only grouping in which they are different is size, where serine is classified as tiny and asparagine is classified as small. This may cause a very small change in the tertiary structure of the protein but probably not one that will be significant.

In those genotypes usually associated with LGV infection at position 5024 there was a single nucleotide substitution from adenine to cytosine that resulted in an amino acid change from lysine to threonine.

At position 5112 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from thymine to guanine that results in a change to the amino acid sequence from a serine to arginine. Serine is quite common in protein functional centres. The hydroxyl group is fairly reactive, being able to form hydrogen bonds with a variety of polar substrates. Arginine is quite frequent in protein active or binding sites. The positive charge means that they can interact with negatively charged non-protein atoms. So it is possible that this substitution may have a functional role.

In those genotypes usually associated with ocular infection at position 5114 there was a single nucleotide substitution from adenine to guanine that resulted in an amino acid change from asparagine to serine. Due to the similar properties of serine and asparagine it is unlikely that there will be a significant change to the tertiary structure of the protein as both are polar, non-charged amino acids. The only grouping in which they are different is size, where serine is classified as tiny and asparagine is classified as small. This may cause a very small change in the tertiary structure of the protein but probably not one that will be significant.

At position 5270 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from guanine to adenine that results in a change to the amino acid sequence from a serine to tyrosine. Tyrosine is classified as aromatic and hydrophobic and thus likes to be contained in a hydrophobic core of a protein. Serine is classified as tiny and slightly polar. A common role for tyrosine and serine within intracellular proteins is phosphorylation it is possible that this is the case here.

In those genotypes usually associated with genital tract infection at position 5299 there was a single nucleotide substitution from guanine to adenine that resulted in an amino acid change from isoleucine to valine. Both amino acids have similar properties; they are both classified as aliphatic and hydrophobic.

Both also have C-beta branched side chains. This means that there is a lot more bulkiness near to the protein backbone, thus meaning that these amino acids are more restricted in the conformations the main - chain can adopt. it is more difficult therefore for these amino acids to adopt an alpha-helical conformation, though it is easy for them to be within beta pleated sheets.

In those genotypes usually associated with genital tract infection at position 5328 there was a single nucleotide substitution from cytosine to thymine that resulted in an amino acid change from arginine to serine. As noted before this may have a functional role as both amino acids have known reactive side chains.

At position 5333 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from cytosine to thymine that results in a change to the amino acid sequence from a threonine to isoleucine as stated before both of these amino acids are C - beta branched and so this may be a structurally dependant change in the sequence rather than a functionally dependant change.

There is one intergenic point mutation at position 5540. This is a substitution from cytosine to adenine in those genotype usually associated with LGV infection. Because of the intergenic nature of the substitution there was no resulting amino acid change.

7.9 ORF 6

Overall 1 difference was noted in the 10 aligned sequences of ORF 6 this was a substitution.

At position 5577 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from adenine to cytosine that results in a change to the amino acid sequence from an arginine to serine. As noted before this may have a functional role as both amino acids have known reactive side chains.

There is one intergenic point mutation at position 5873. This is a substitution from cytosine to thymine in those genotype usually associated with LGV infection. Because of the intergenic nature of the substitution there was no resulting amino acid change.

7.10 ORF 7

Overall four different changes in the 10 aligned sequences were noted in ORF 7 these were all substitutions.

In those genotypes usually associated with genital tract infection at position 6064 there was a single nucleotide substitution from adenine to guanine that resulted in an amino acid change from asparagine to histidine. It is rare that histidine is exchanged for other amino acids so this change may be significant although neither of these two amino acids have known interchangeable functions, and thus the function of this change cannot be hypothesised.

At position 6078 in those genotypes usually associated with LGV infection there is a single nucleotide substitution from guanine to adenine that results in a change to the amino acid sequence from phenylalanine to leucine. Phenylalanine is classified as aromatic where as leucine is considered aliphatic both however, are hydrophobic and prefer to reside in a hydrophobic core of the tertiary protein structure.

In those genotypes usually associated with genital tract infection at position 6167 there was a single nucleotide substitution from thymine to cytosine that resulted in an amino acid change from leucine to serine. As these amino acids have no common traits the hypothesis of this substitution is unknown.

In those genotypes usually associated with ocular infection at position 6335 there was a single nucleotide substitution from cytosine to thymine that resulted in an amino acid change from alanine to valine. Both the side chains of valine and alanine are non-reactive, and are thus rarely directly involved in protein function, though they can play a role in substrate recognition. Also both amino acids are hindered in their conformations due to the fact that they both contain C-beta carbons in there formations. It is hypothesised therefore that this substitution may be due to the structure of the tertiary protein.

7.11 ORF 8

Overall six different changes in the 10 aligned sequences were noted in ORF 8 these were all substitutions.

Table 7.7 shows those changes to the sequences that do no result in a change to the translated sequences

Table 7.7 Changes in ORF 8 that do not result in a change to the translated sequence

Sequences with the bp change	Type of change	Position	Bp change
L1, L2 and L2 LON	Substitution	6750	C - T
A, CTA, CTB, and JALI	Substitution	6767	G - A
D, SW2 and SW3	Substitution	6773	T - C
L1, L2 and L2 LON	Substitution	6829	G - A

In those genotypes usually associated with LGV infection at position 6714 there was a single nucleotide substitution from guanine to adenine that resulted in an amino acid change from aspartate to asparagine. Asparagine and aspartate are polar amino acids. Aspartate, contains an oxygen in place of the amino group in Asparagine however, this is the only difference between the amino acids. Therefore because of their similarities it is hypothesised that the resulting changes to the protein may not be considered significant.

At position 7255 in those genotypes usually associated with ocular infection there is a single nucleotide substitution from guanine to adenine that results in a change to the amino acid sequence from methionine to isoleucine. Both the isoleucine and methionine side chains are very non-reactive, and are thus rarely directly involved in protein function, though they can play a role in substrate recognition. This may be the role in this case

7.12 Intergenic point mutations

The following point mutations were noted in the intergenic sequence. These mutations fall outside of the open reading frames and so have no effect on the translated protein.

At position 7457 there is an adenine and thymine 2 bp insertion in those genotypes usually associated with genital tract infection.

In those genotypes usually associated with ocular infection at position 7458 there is a single nucleotide substitution from cytosine to thymine.

At position 7472 there is a 2bp insertion of thymine in those genotypes usually associated with ocular infection.

In those genotypes usually associated with genital infection at position 7489 there is a single nucleotide substitution from thymine to cytosine.

Lastly there is a four bp insertion in the Jali serovar B sequence starting at position 7500 that reads guanine cytosine adenine and cytosine.

7.13 Plasmid sequence analysis of unique features

Comparative data analysis could be used to develop chlamydial typing systems for the plasmids as each is unique (see table 7.8) and this might generate a secondary marker but, it is unknown if plasmids evolve with their chlamydial hosts and the plasmid is probably not the best choice for this because plasmid free strains of *C.trachomatis* although rare have been described.

Table 7.8 Examples of unique features in the ORF1 region of the plasmid compared to CTB genotype

Sequence	Unique feature
A	8bp insertion at position 80
CTA	One 22bp repeat missing
Jali	A at position 805
D	A at position 120
SW2	377bp deletion at position 653
SW3	C at position 910
L1	T at position 910
L2	Single bp deletion at position 910
L2b	Features of L2 apart from the single bp deletion

CHAPTER 8 GENERAL DISCUSSION AND FUTURE WORK

8.0 Jersey *OmpA* project

At present there has been the successful development of an *ompA* specific PCR that will amplify o*mpA* from both laboratory cultivated *C.trachomatis* isolates and clinical samples. The successful amplified *ompA* PCR products came from a range of different sites of infection these included eye swabbings and male urethral swabs. However, most of the data was collected from endocervical swabs. A major concern because of the selective population of Jersey and the fact that it is a small island was that the Chlamydia samples on Jersey were all the same genotype or of very limited variability. This null hypothesis that there was a single *ompA* genotype distribution in Jersey, has been deemed void by the fact that we have identified five 'genotypes' from the BD Probe Tech *C.trachomatis* positive samples. The genotypes detected correspond to 'serovars' D, E, G, J and K. The most prevalent genotype found corresponded to genotype E this was in accordance with other data published in the Chlamydial field (Millman *et al.,* 2001)

Although the PCR developed was successful in identifying the *ompA* gene from *C.trachomatis* positive samples, of the 341 BD Probe Tech *C.trachomatis* positive samples screened, 38 were PCR *ompA* positive. However, some of these positives were not of a suitable quality or quantity to be genotyped via sequencing or the samples were lost during the gel extraction stage of the protocol. This meant that only 23 of the original 341 BD Probe Tech *C.trachomatis* positive samples were genotyped. The rate of *ompA* PCR positivity from the BD Probe Tech *C.trachomatis* positive samples was extremely low and with such a small 'genotyped' sample set there was not enough data to make valid and reliable conclusions from any resulting analyses. Accordingly *C.trachomatis* positive samples from Southampton were collected to

provide firstly a sample set that was sizable enough to draw meaningful conclusions and secondly, to give a genotypically comparative mainland UK population with which the Jersey's island population could be analysed.

The *C.trachomatis* positive samples collected in Southampton until the summer of 2007 were screened using the Micro Trak II *Chlamydia* EIA kit (Trinity Biotech plc.). In total 152 of these *C.trachomatis* positive samples were collected and were stored at -80^{0}C ready for *ompA* screening. One of the samples collected from the Southampton region has been provisionally tested with the CT1 and CT5 primers and yielded promising *ompA* positive result when amplified using the conditions described by Ngandjio *et al.,* (2003). It is hypothesised that these samples will give us a small selection of serovars similar to those presenting in Jersey. In addition it is hypothesised that the most prevalent genotype in the Southampton area is that of genotype E as this is the most prevalent serovar found elsewhere in the field (Millman *et al.,* 2001).

For scientific rigour there is a need to guard against the use of a single genetic marker which may possess unforeseen properties such as possible gene mobility as reported by Gomes *et al.,* (2004) that would bias the survey. A considerable amount of work has been completed with the *ompA* of *C.trachomatis,* however, it is a large gene to amplify and is only a single genotypic marker. Even though *ompA* appears to be a genetic marker that truly reflects clonal stability it would be beneficial to have a secondary marker to confirm serovar identity of the MOMP as there is evidence for gene recombination close to *ompA* (Millman *et al.,* 2001). At the start of this project only three *C.trachomatis* genome sequences were available. During my work and training at the Sanger Centre I assisted with the analysis and genome annotation of two further genome sequences. From the preliminary genome data analysis and literature research conducted, suitable secondary potential gene markers with viability include the Translocated Actin Recruiting

Phosphoprotein (TARP). Other potential secondary gene marker candidates included the PMPs or putative membrane proteins originally described by (Longbottom *et al.,* 1998). Another candidate marker sequence is the cytotoxin locus, which exhibits extensive mutations and deletions among different serovars (Carlson *et al.,* 2004). Furthermore, Carlson et al., (2004), found sequence motifs in the cytotoxin locus specific to oculotropic, urogenitropic and invasive serovars (Carlson *et al.,* 2004), which means information from the study of this locus during this investigation, should reinforce the evidence base relating *ompA* typing to disease presentations.

To date primers have been designed for both the TARP and PMP potential secondary markers. Primer design for both potential secondary markers was based around the A, D and L2 genomes already published. As a starting point the TARP genes were located in all three genomes. As not all of the genomes have been annotated fully this required in some cases searching through the sequence using the ACT alignment programme for the CT reference number that the Sanger allocates to ORFs. This requires searching the sequence for patterns and identity between the aligned sequences. Once located all three TARP genes were analysed for differences between them and the 50 bps of sequence around each TARP gene was mapped. Again using the alignment package ACT, a forward and a reverse primer were designed to amplify the TARP gene. Each primer was designed to anneal to a small piece of sequence that was identical in all three genomes. This was also the way in which the primers were designed for the PMPs. Table 8.1 shows the sequences for each of the designed primers

Table 8.1 TARP and PMP primer sequences

Primer Name	Primer Sequence 5'-3'
TARP forward	TTG TCG ATT GTG ATG AGG AGA CA
TARP reverse	ACG TTA CCC GGT ATG CTG TTA T
PMP F1	GAA TGG CTC CGC CTT CTC TTA TTT T
PMP F2	CTT ATC TTC AGC GCA TTC GTC CTT C
F-3	ATT CTA GAA GGA ACC TCT CT
F-4	TGA CAG GGA ATC GAA CT CT
F-5	GAC TGA AAT TTC CCT TTA GA
F-6	AAA TTA TAT CCA TAC AGG GA
F-7	GCA TTC GTC CTT CTC TCT TT

On completion of the selected secondary marker research to optimise the PCR reactions, gene screening of the secondary marker is planned using the *ompA* genotyped BD Probe Tech *C.trachomatis* positive samples collected from Jersey and the Roche Cobas Taqman *C.trachomatis* positive samples collected from Southampton. Sequencing of the BD Probe Tech *C.trachomatis* positive samples collected from Jersey and the Roche Cobas Taqman *C.trachomatis* positive samples collected from Southampton positive for the chosen secondary marker will be completed to facilitate complete genotype linkage analysis. This will be completed to ensure that the two selected genes show identity in the resulting genotype of each *C.trachomatis* positive Jersey clinical sample. The genotype distribution of the BD Probe Tech *C.trachomatis* samples collected from Jersey and the Roche Cobas Amplicor *C.trachomatis* positive samples collected from Southampton will then be related to the published genomes to see if any cross linkage studies can draw genotype specific conclusions to the biovar or sero-grouping (B, C and intermediate). The data from the secondary marker PCR analysed concurrently with *ompA,* for serovar identification will aid an investigation and comparative analysis of ocular, genital tract and LGV strain infections. This will also involve the comparative analysis of the A, D, and L2 published genomes

The collection of C.trachomatis positive samples from Jersey and Southampton with different detection technologies offers the ideal approach to screen the new variant *C.trachomatis* from Malmo, Sweden.

Adams, E. J., Charlett, A., Edmunds, W. J. & Hughes, G. (2004). Chlamydia trachomatis in the United Kingdom: a systematic review and analysis of prevalence studies. *Sex Transm Infect* **80**, 354-362.

Bedson, S.P. and Bland J.O.W. (1934). The developmental forms of Psittacosis Virus. *Brit. J. Exptl, Path.* **15**, 243-247.

Brown, A. E., Sadler, K. E., Tomkins, S. E. & McGarrigle, C., LaMontagne, D., Goldberg, D., Tookey, P., Smyth, B., Thomas, D., Murphy, G., Parry, J., Evans, B., Gill, O., Ncube, F.,and Fenton,K.,(2004). Recent trends in HIV and other STIs in the United Kingdom: data to the end of 2002. *Sex Transm Infect* **80**, 159-166.

Caldwell, H.D., Kromhout, J.; Schachter, J.,(1981). Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. *Infect Immun.* Mar;**31**(3), 1161–1176.

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

Cates, W., Jr. (1999). Estimates of the incidence and prevalence of sexually transmitted diseases in the United States. American Social Health Association Panel. *Sex Transm Dis* **26**, S2-S7.

Collier L.H, Duke-Elder S & Jones, B. R. (1958). Experimental trachoma produced by cultured virus. *British Journal of Ophthalmology* **42**, 705-720.

Collier L.H & Sowa J (1958). Isolation of trachoma virus in embryonate eggs. *Lancet* **1**, 993-996.

Cox H.R (1938). Use of yolk sac of developing chick embryo as medium for growing Rickettsiae of Rocky Mountain spotted fever and typhus groups. 53 edn, pp. 2241-2247.

Cox, M. M. (1998).A broadening view of recombinational DNA repair in bacteria. *Genes Cells* **3**, 65-78.

Darougar, S., Forsey, T., Brewerton, D. A. & Rogers, K. L. (1980). Prevalence of anti-Chlamydial antibody in London blood donors. *Br J Verner Dis* **56**, 404-407.

Dunlop, E. M. C., Jones, B. R. & Al-Hussaini, M. K. (1964). Genital infection in association with TRIC virus infection of the eye. III. Clinical and other findings. Preliminary report. *Br J Vener Dis* **40**, 33-42.

Dwyer, R. S., Treharne, J. D., Jones, B. R. & Herring, J. (1972). Chlamydial infection. Results of micro-immunofluorescence tests for the detection of type-specific antibody in certain Chlamydial infections. *Br J Vener Dis* **48**, 452-459.

Ejere, H., Alhassan, M. B. & Rabiu, M. (2004). Face washing promotion for preventing active trachoma. *Cochrane Database Syst Rev*, CD003659.

Euzéby, J. P. (1980). Approved lists of bacterial names. *Med J Aust* **2**, 3-4.

Everett, K. D. E., Bush, R. M. & 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 for the identification of organisms. *Int J Syst Bacteriol* **49**, 415-440.

Fenton, K. A., Korovessis, C., Johnson, A. M. & other authors (2001). Sexual behaviour in Britain: reported sexually transmitted infections and prevalent genital Chlamydia trachomatis infection. *Lancet* **358**, 1851-1854.

Frutos, R., Federici, B. A., Revet, B. & Bergoin, M. (1994). Taxonomic studies of Rickettsiella, Rickettsia, and Chlamydia using genomic DNA. *J Invertebr Pathol* **63**, 294-300.

Fukushi, H. & Hirai, K. (1993).*Chlamydia pecorum* - The fourth species of genus *Chlamydia*. *Microbiology & Immunology* **37**, 515-522.

Geisler, W. M., Tang, J., Wang, C., Wilson, C. M. & Kaslow, R. A. (2004). Epidemiological and genetic correlates of incident Chlamydia trachomatis infection in North American adolescents. *J Infect Dis* **190**, 1723-1729.

Gerbase, A. C., Rowley, J. T. & Mertens, T. E. (1998). Global epidemiology of sexually transmitted diseases. *Lancet* **351 Suppl 3**, 2-4.

Golden, M. R., Schillinger, J. A., Markowitz, L. & St Louis, M. E. (2000). Duration of untreated genital infections with chlamydia trachomatis: a review of the literature. *Sex Transm Dis* **27**, 329-337.

Gordon, F. B. & Quan, A. L. (1965) Isolation of the trachoma agent in cell culture. *Proc Soc Exp Biol Med.* Feb;**118**, 354–359.

Gordon, F. B. & Quan, A. L. (1971). Isolation of *Chlamydia trachomatis* from the human genital tract by cell culture: a summery In:Trachoma and Related Disorders. pp. 476-487. Edited by Nichols, R. L.

Gotz, H. M., van Bergen, J. E., Veldhuijzen, I. K. & other authors (2006). Lessons learned from a population-based chlamydia screening pilot. *Int J STD AIDS* **17**, 826-830.

Grayston, J. T. (1989).*Chlamydia pneumoniae,* strain TWAR. *Chest* **95**, 664-669.

Grayston, J. T. & Wang, S.-P. (1975). New knowledge of chlamydiae and the diseases they cause. *Infect Agents Dis* **132**, 87-105.

Halse, T. A., Musser, K. A. & Limberger, R. J. (2006).A multiplexed real-time PCR assay for rapid detection of Chlamydia trachomatis and

identification of serovar L-2, the major cause of Lymphogranuloma venereum in New York. *Mol Cell Probes* **20**, 290-297.

Harkness, A. H., (1950) Non-gonococcal urethritis; a review of a monograph by Dr. A. H. Harkness *Med Illus.* Sep;4(9), 471-4.

Harper, I. A., Dwyer, R. s. C., Garland, J. A., Jones, B. R., Treharne, J. D., Dunlop, E. M. C., Freedman, A. & Race, J. W. (1967). Infection by Bedsoniae and the possiblilty of spurious isolation: 1 cross infection of eggs during culture. *Am J Ophthalmol* **63**, 1064-1073.

Hayes, L. J., Yearsley, P., Treharne, J. D., Ballard, R. A., Fehler, G. H. & Ward, M. E. (1994). Evidence for naturally occurring recombination in the gene encoding the major outer membrane protein of lymphogranuloma venereum isolates of *Chlamydia trachomatis*. *Infect Immun* **62**, 5659-5663.

Herrmann, B. & Egger, M. (1995). Genital Chlamydia trachomatis infections in Uppsala County, Sweden, 1985-1993: declining rates for how much longer? *Sex Transm Dis* **22**, 253-260.

Hsu, M. C., Tsai, P. Y., Chen, K. T., Li, L. H., Chiang, C. C., Tsai, J. J., Ke, L. Y., Chen, H. Y. & Li, S. Y. (2006). Genotyping of Chlamydia trachomatis from clinical specimens in Taiwan. *J Med Microbiol* **55**, 301-308.

Jones, B. R., Collier, L. H. & Smith, C. H. (1959). Isolation of virus from inclusion blennorrhoea. *Lancet* **1**, 902-905.

Jones, H., Rake, G. & Stearns, B. (1945). Studies on L.G.V. III. *J Infect Dis* **76**, 55-69.

Jurstrand, M., Falk, L., Fredlund, H., Lindberg, M., Olcen, P., Andersson, S., Persson, K., Albert, J. & Backman, A. (2001). Characterization of Chlamydia trachomatis omp1 genotypes among sexually transmitted disease patients in Sweden. *J Clin Microbiol* **39**, 3915-3919.

Kaltenboeck, B., Kousoulas, K. G. & Storz, J. (1993). Structures of and allelic diversity and relationships among the major outer-membrane protein (ompA) genes of the four Chlamydial species. *J Bacteriol* **175**, 487-502.

Kamwendo, F., Forslin, L., Bodin, L. & Danielsson, D. (1998). Programmes to reduce pelvic inflammatory disease--the Swedish experience. *Lancet* **351 Suppl 3**, 25-28.

Kurnosova, L.M., Lenkevich, M.M., (1964). Mode of action of sulfonamides on trachoma virus.Acta Virol. Jul;8:350-8

Longbottom, D., Russell, M., Dunbar, S. M., Jones, G. E. & Herring, A. J. (1998). Molecular cloning and characterization of the genes coding for the highly immunogenic cluster of 90-kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. *Infect Immun* **66**, 1317-1324.

Low, N. & Egger, M. (2002). What should we do about screening for genital chlamydia? *Int J Epidemiol* **31**, 891-893.

Mabey, D. & Peeling, R. W. (2002). Lymphogranuloma venereum. *Sex Transm Infect* **78**, 90-92.

Macleod, J., Salisbury, C., Low, N. & other authors (2005). Coverage and uptake of systematic postal screening for genital Chlamydia trachomatis and prevalence of infection in the United Kingdom general population: cross sectional study. *BMJ* **330**, 940.

McKay, L., Clery, H., Carrick-Anderson, K., Hollis, S. & Scott, G. (2003). Genital Chlamydia trachomatis infection in a subgroup of young men in the UK. *Lancet* **361**, 1792.

Millman, K. L., Tavare, S. & Dean, D. (2001). Recombination in the ompA gene but not the omcB gene of Chlamydia contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. *J Bacteriol* **183**, 5997-6008.

Molano, M., Meijer, C. J., Morre, S. A., Pol, R. & van den Brule, A. J. (2004). Combination of PCR targeting the VD2 of omp1 and reverse line blot analysis for typing of urogenital Chlamydia trachomatis serovars in cervical scrape specimens. *J Clin Microbiol* **42**, 2935-2939.

Ngandjio, A., Clerc, M., Fonkoua, M. C. & other authors (2003). Screening of volunteer students in Yaounde (Cameroon, Central Africa) for Chlamydia trachomatis infection and genotyping of isolated C. trachomatis strains. *J Clin Microbiol* **41**, 4404-4407.

Noone, A., Spiers, A., Allardice, G., Carr, S., Flett, G., Brown, A. & Twaddle, S. (2004). Opportunistic screening for genital Chlamydia trachomatis infection and partner follow-up in family planning clinics in three Scottish cities. *J Fam Plann Reprod Health Care* **30**, 84-85.

Oriel, J. D. & Ridgway, G. L. (1982).*Genital Infection by Chlamydia trachomatis*. Edward Arnold Ltd.

Page L.A (1966). Revision of the family Chlamydiaceae (Rake). *Int J Syst Bacteriol* **16**, 223-252.

Page L.A (1968). Proposal for the recognition of two species in the genus Chlamydia. *Int J Syst Bacteriol* **18**, 51-66.

Pudjiatmoko, Fukushi, H., OChiai, Y., Yamaguchi, T. & Hirai, K. (1997). Phylogenetic analysis of the genus *Chlamydia* based on 16S rRNA gene sequences. *Int J Syst Bacteriol* **47**, 425-431.

Rabiu, M., Alhassan, M. & Ejere, H. (2005). Environmental sanitary interventions for preventing active trachoma. *CoChrane Database Syst Rev*, CD004003.

Reed, S. I., Anderson, L. E. & Jenkin, H. M. (1981). Use of cycloheximide to study independent lipid metabolism of Chlamydia trachomatis cultivated in mouse L cells grown in serum-free medium. *Infect Immun* **31**, 668-673.

Rein, D. B., Kassler, W. J., Irwin, K. L. & Rabiee, L. (2000). Direct medical cost of pelvic inflammatory disease and its sequelae: decreasing, but still substantial. *Obstet Gynecol* **95**, 397-402.

Philip, R. N., D. A Hill, A. B. Greaves, F. B. Gordon, A. L. Quan, R. K. Gerloff, and L. A. Thomas (1971). Study of Chlamydia in patients with lymphogranuloma venereum and urethritis attending a veneral diseases clinic. Brit. *J. Vener. Dis.* 47:114-121.

Ripa, T. & Nilsson, P. (2006).A variant of Chlamydia trachomatis with deletion in cryptic plasmid: implications for use of PCR diagnostic tests. *Euro Surveill* **11**, E061109.

Rogers, D. G. & Andersen, A. A. (1996). Intestinal lesions caused by two swine chlamydial isolates in gnotobiotic pigs. *J Vet Diagn Invest* **8**, 433-440.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977).DNA sequencing with chain-terminating inhibitors. *Biochemistry (Mosc)* **74**, 5463-5467.

Schachter, J. (1990).chlamydiae. In *Microbiolgy*, 4th edn, p. 699. Edited by Davis B.D., Dulbecco R., Herman N & Eisen: Lippincott.

Schachter, J. (1999). We must be realistic in evaluating rapid diagnostic tests. *Sex Transm Dis* **26**, 241-242.

Schachter, J., Christiansen, G., Clarke, I. N. & other authors (2002). Chlamydial Infections Proceedings of the Tenth International Symposium on Human Chlamydial Infections.

Schleifer, K. H. & Stackebrandt, E. (1983). Molecular systematics of prokaryotes. *Annu Rev Microbiol* **37**, 143-187.

Spaargaren, J., Fennema, H. S., Morre, S. A., de Vries, H. J. & Coutinho, R. A. (2005). New lymphogranuloma venereum Chlamydia trachomatis variant, Amsterdam. *Emerg Infect Dis* **11**, 1090-1092.

Stephens, R.S., Mullenbach, G., Sanches-Pescador, R. & Agabian, N. (1986) Sequence analysis of major outer membrane protein gene from Chlamydia trachomatis serovar L2. *J. Bactriol.*, 168, 1277-1282.

Stephens, R. S., Sanchez-Pescador, M., Wagar, E. A., Inouye, C. & Urdea, M. S. (1987). Diversity of *Chlamydia trachomatis* Major Outer Membrane Protein Genes. *J Bacteriol* **169**, 3879-3885.

Stephens, R. S., Tam, M. R., Kuo, C.-C. & Nowinski, R. C. (1982). Monoclonal antibodies to *Chlamydia trachomatis*: antibody specificities and antigen characterization. *J Immunol* **128**, 1083-1089.

T'ang, F. F., Chang, H. L., Huang, Y. T. & Wang K.C (1957). Studies on the aetiology of trachoma with special reference to isolation of the virus in chick embryo. *Chin Med J (Engl)* **75**, 429-447.

Takahashi, T., Masuda, M., Tsuruno, T., Mori, Y., Takashima, I., Hiramune, T. & Kikuchi, N. (1997). Phylogenetic analyses of *Chlamydia psittaci* strains from birds based on 16S rRNA gene sequence. *J Clin Microbiol* **35**, 2908-2914.

Thomas, N. S. (1995). Ph.D. Thesis. University of Southampton, U.K.

Thomson, N.R., Holden, M.T., Carder, C., Lennard, N., Lockey, S.J., Marsh, P., Skipp, P., O'Connor, C.D., Goodhead, I., Norbertzcak, H., Harris, B., Ormond, D., Rance, R., Quail, M.A., Parkhill, J., Stephens, R.S., Clarke, I.N., (2008) Chlamydia trachomatis: Genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res* **18**:161-71.

Thygeson P & Mengert W.F. (1936). The virus of inclusion conjunctivitis. Further observations. *Archives of Opthalmology* **15**, 377-410.

Thygeson P & Stone W (1942). Epidemiolgy of inclusion conjunctivitis. *Arch Ophthalmol* **27**, 91-122.

Van der Bij, A. K., Spaargaren, J., Morre, S. A., Fennema, H. S., Mindel, A., Coutinho, R. A. & de Vries, H. J. (2006). Diagnostic and clinical implications of anorectal lymphogranuloma venereum in men who have sex with men: a retrospective case-control study. *Clin Infect Dis* **42**, 186-194.

Wang, S. P., Kuo, C. C., Barnes, R. C., Stephens, R. S. & Grayston, J. T. (1985). Immunotyping of Chlamydia trachomatis with monoclonal antibodies. *J Infect Dis* **152**, 791-800.

Weisburg, W. G., Hatch, T. P. & Woese, C. R. (1986). Eubacterial origin of Chlamydia. *J Bacteriol* **167**, 570-574.

Yuan, Y., Zhang, Y.-X., Watkins, N. G. & Caldwell, H. D. (1989). Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. *Infect Immun* **57**, 1040-1049.

Web References

[1] http://data.euro.who.int/cisid/?TabID=143163 (02/010/2007)

sequence	orf	position	type of change	bp change	Pro change
cp000052	int				no
CTA	int	1 to 22	deletion 22bp	TTTGCAACTCTTGGTGGTAGAC	no
CTB	int				no
JALI	int				no
NC_001372	int				no
SW2	int				no
SW3	int				no
X06707	int				no
X07547	int				no
L2 LON	int				no
cp000052	int	80 to 88	insert 8bp	GTGGTAGG	no
CTA	int				no
CTB	int				no
JALI	int				no
NC_001372	int				no
SW2	int				no
SW3	int				no
X06707	int				no
X07547	int				no
L2 LON	int				no
cp000052	int	111	substitution	A	no
CTA	int	111	substitution	A	no
CTB	int	111	substitution	A	no
JALI	int	111	substitution	A	no
NC_001372	int	111	substitution	A	no
SW2	int	111	substitution	A	no
SW3	int	111	substitution	A	no
X06707	int	111	substitution	G	no
X07547	int	111	substitution	G	no
L2 LON	int	111	substitution	G	no
cp000052	int	120	substitution	C	no
CTA	int	120	substitution	C	no
CTB	int	120	substitution	C	no
JALI	int	120	substitution	C	no
NC_001372	int	120	substitution	A	no
SW2	int	120	substitution	C	no
SW3	int	120	substitution	C	no
X06707	int	120	substitution	C	no
X07547	int	120	substitution	C	no
L2 LON	int	120	substitution	C	no

orf 1 165 -					
cp000052	ORF 1	194	substitution	A	no
CTA	ORF 1	194	substitution	C	no
CTB	ORF 1	194	substitution	C	no
JALI	ORF 1	194	substitution	C	no
NC_001372	ORF 1	194	substitution	A	no
SW2	ORF 1	194	substitution	A	no
SW3	ORF 1	194	substitution	A	no
X06707	ORF 1	194	substitution	A	no
X07547	ORF 1	194	substitution	A	no
L2 LON	ORF 1	194	substitution	A	no
cp000052	ORF 1	200	substitution	C	no
CTA	ORF 1	200	substitution	C	no
CTB	ORF 1	200	substitution	C	no
JALI	ORF 1	200	substitution	C	no
NC_001372	ORF 1	200	substitution	T	no
SW2	ORF 1	200	substitution	T	no
SW3	ORF 1	200	substitution	T	no
X06707	ORF 1	200	substitution	C	no
X07547	ORF 1	200	substitution	C	no
L2 LON	ORF 1	200	substitution	C	no
cp000052	ORF 1	246	substitution	C	his
CTA	ORF 1	246	substitution	C	his
CTB	ORF 1	246	substitution	C	his
JALI	ORF 1	246	substitution	C	his
NC_001372	ORF 1	246	substitution	T	tyr
SW2	ORF 1	246	substitution	T	tyr
SW3	ORF 1	246	substitution	T	tyr
X06707	ORF 1	246	substitution	C	his
X07547	ORF 1	246	substitution	C	his
L2 LON	ORF 1	246	substitution	C	his
cp000052	ORF 1	398	substitution	C	no
CTA	ORF 1	398	substitution	C	no
CTB	ORF 1	398	substitution	C	no
JALI	ORF 1	398	substitution	C	no
NC_001372	ORF 1	398	substitution	T	no
SW2	ORF 1	398	substitution	T	no
SW3	ORF 1	398	substitution	T	no
X06707	ORF 1	398	substitution	C	no
X07547	ORF 1	398	substitution	C	no
L2 LON	ORF 1	398	substitution	C	no

cp000052	ORF 1	410	substitution	C	no
CTA	ORF 1	410	substitution	C	no
CTB	ORF 1	410	substitution	C	no
JALI	ORF 1	410	substitution	C	no
NC_001372	ORF 1	410	substitution	A	no
SW2	ORF 1	410	substitution	A	no
SW3	ORF 1	410	substitution	A	no
X06707	ORF 1	410	substitution	C	no
X07547	ORF 1	410	substitution	C	no
L2 LON	ORF 1	410	substitution	C	no

cp000052	ORF 1				
CTA	ORF 1				
CTB	ORF 1				
JALI	ORF 1				
NC_001372	ORF 1				
SW2	ORF 1	653- 1030	deletion 377bp		
SW3	ORF 1				
X06707	ORF 1				
X07547	ORF 1				
L2 LON	ORF 1				

cp000052	ORF 1	785	substitution	A	no
CTA	ORF 1	785	substitution	A	no
CTB	ORF 1	785	substitution	A	no
JALI	ORF 1	785	substitution	A	no
NC_001372	ORF 1	785	substitution	A	no
SW2	ORF 1	785	substitution		
SW3	ORF 1	785	substitution	A	no
X06707	ORF 1	785	substitution	G	no
X07547	ORF 1	785	substitution	G	no
L2 LON	ORF 1	785	substitution	G	no

cp000052	ORF 1	805	substitution	G	cys
CTA	ORF 1	805	substitution	G	cys
CTB	ORF 1	805	substitution	G	cys
JALI	ORF 1	805	substitution	A	tyr
NC_001372	ORF 1	805	substitution	G	cys
SW2	ORF 1	805	substitution		
SW3	ORF 1	805	substitution	G	cys
X06707	ORF 1	805	substitution	G	cys
X07547	ORF 1	805	substitution	G	cys
L2 LON	ORF 1	805	substitution	G	cys

cp000052	ORF 1	894	substitution	C	his
CTA	ORF 1	894	substitution	C	his
CTB	ORF 1	894	substitution	C	his
JALI	ORF 1	894	substitution	C	his
NC_001372	ORF 1	894	substitution	C	his
SW2	ORF 1	894	substitution		
SW3	ORF 1	894	substitution	C	his
X06707	ORF 1	894	substitution	T	tyr
X07547	ORF 1	894	substitution	T	tyr
L2 LON	ORF 1	894	substitution	T	tyr
cp000052	ORF 1				
CTA	ORF 1				
CTB	ORF 1				
JALI	ORF 1				
NC_001372	ORF 1				
SW2	ORF 1				
SW3	ORF 1				
X06707	ORF 1				
X07547	ORF 1	910	deleltion 1bp	A	premature termination
L2 LON	ORF 1	910	deleltion 1bp	A	premature termination
cp000052	ORF 1	923	substitution	C	no
CTA	ORF 1	923	substitution	C	no
CTB	ORF 1	923	substitution	C	no
JALI	ORF 1	923	substitution	C	no
NC_001372	ORF 1	923	substitution	C	no
SW2	ORF 1	923	substitution		
SW3	ORF 1	923	substitution	C	no
X06707	ORF 1	923	substitution	T	no
X07547	ORF 1	923	substitution	T	no
L2 LON	ORF 1	923	substitution	T	no
cp000052	ORF 1				
CTA	ORF 1				
CTB	ORF 1				
JALI	ORF 1				
NC_001372	ORF 1				
SW2	ORF 1				
SW3	ORF 1				
X06707	ORF 1				
X07547	ORF 1	948	END		premature termination
L2 LON	ORF 1	948	END		premature termination

cp000052	ORF 1	1004	substitution	C	no
CTA	ORF 1	1004	substitution	C	no
CTB	ORF 1	1004	substitution	C	no
JALI	ORF 1	1004	substitution	C	no
NC_001372	ORF 1	1004	substitution	T	no
SW2	ORF 1	1004	substitution		
SW3	ORF 1	1004	substitution	T	no
X06707	ORF 1	1004	substitution	T	no
X07547	INT	1004	substitution	T	no
L2 LON	INT	1004	substitution	T	no
cp000052	ORF 1	1051	substitution	G	ser
CTA	ORF 1	1051	substitution	G	ser
CTB	ORF 1	1051	substitution	G	ser
JALI	ORF 1	1051	substitution	G	ser
NC_001372	ORF 1	1051	substitution	T	Ile
SW2	ORF 1	1051	substitution	T	Ile
SW3	ORF 1	1051	substitution	T	Ile
X06707	ORF 1	1051	substitution	G	ser
X07547	INT	1051	substitution	G	no
L2 LON	INT	1051	substitution	G	no
cp000052	ORF 1	1071	substitution	A	leu
CTA	ORF 1	1071	substitution	A	leu
CTB	ORF 1	1071	substitution	A	leu
JALI	ORF 1	1071	substitution	A	leu
NC_001372	ORF 1	1071	substitution	A	leu
SW2	ORF 1	1071	substitution	A	leu
SW3	ORF 1	1071	substitution	A	leu
X06707	ORF 1	1071	substitution	C	pro
X07547	INT	1071	substitution	C	no
L2 LON	INT	1071	substitution	C	no
cp000052	ORF 1	1082	END		
CTA	ORF 1	1082	END		
CTB	ORF 1	1082	END		
JALI	ORF 1	1082	END		
NC_001372	ORF 1	1082	END		
SW2	ORF 1	1082	END		
SW3	ORF 1	1082	END		
X06707	ORF 1	1082	END		
X07547	INT				no
L2 LON	INT				no

cp000052	INT	1126	substitution	C	no
CTA	INT	1126	substitution	C	no
CTB	INT	1126	substitution	C	no
JALI	INT	1126	substitution	C	no
NC_001372	INT	1126	substitution	C	no
SW2	INT	1126	substitution	C	no
SW3	INT	1126	substitution	C	no
X06707	INT	1126	substitution	T	no
X07547	INT	1126	substitution	T	no
L2 LON	INT	1126	substitution	T	no

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF 2	1132	END		
CTA	ORF 2	1132	END		
CTB	ORF 2	1132	END		
JALI	ORF 2	1132	END		
NC_001372	ORF 2	1132	END		
SW2	ORF 2	1132	END		
SW3	ORF 2	1132	END		
X06707	ORF 2	1132	END		
X07547	ORF 2	1132	END		
L2 LON	ORF 2	1132	END		
cp000052	ORF 2	1147	substitution	T	leu
CTA	ORF 2	1147	substitution	T	leu
CTB	ORF 2	1147	substitution	T	leu
JALI	ORF 2	1147	substitution	T	leu
NC_001372	ORF 2	1147	substitution	T	leu
SW2	ORF 2	1147	substitution	G	met
SW3	ORF 2	1147	substitution	G	met
X06707	ORF 2	1147	substitution	G	met
X07547	ORF 2	1147	substitution	G	met
L2 LON	ORF 2	1147	substitution	G	met
cp000052	ORF 2	1160	substitution	A	lys
CTA	ORF 2	1160	substitution	G	asn
CTB	ORF 2	1160	substitution	G	asn
JALI	ORF 2	1160	substitution	G	asn
NC_001372	ORF 2	1160	substitution	G	asn
SW2	ORF 2	1160	substitution	G	asn
SW3	ORF 2	1160	substitution	G	asn
X06707	ORF 2	1160	substitution	G	asn
X07547	ORF 2	1160	substitution	G	asn
L2 LON	ORF 2	1160	substitution	G	asn
cp000052	ORF 2	1175	substitution	G	pro
CTA	ORF 2	1175	substitution	G	pro
CTB	ORF 2	1175	substitution	G	pro
JALI	ORF 2	1175	substitution	G	pro
NC_001372	ORF 2	1175	substitution	A	ser
SW2	ORF 2	1175	substitution	A	ser
SW3	ORF 2	1175	substitution	A	ser
X06707	ORF 2	1175	substitution	A	ser
X07547	ORF 2	1175	substitution	A	ser
L2 LON	ORF 2	1175	substitution	A	ser

cp000052	ORF 2	1217	substitution	C	no
CTA	ORF 2	1217	substitution	C	no
CTB	ORF 2	1217	substitution	C	no
JALI	ORF 2	1217	substitution	C	no
NC_001372	ORF 2	1217	substitution	T	no
SW2	ORF 2	1217	substitution	T	no
SW3	ORF 2	1217	substitution	T	no
X06707	ORF 2	1217	substitution	T	no
X07547	ORF 2	1217	substitution	T	no
L2 LON	ORF 2	1217	substitution	T	no
cp000052	ORF 2	1301	substitution	C	no
CTA	ORF 2	1301	substitution	C	no
CTB	ORF 2	1301	substitution	C	no
JALI	ORF 2	1301	substitution	C	no
NC_001372	ORF 2	1301	substitution	T	no
SW2	ORF 2	1301	substitution	T	no
SW3	ORF 2	1301	substitution	T	no
X06707	ORF 2	1301	substitution	C	no
X07547	ORF 2	1301	substitution	C	no
L2 LON	ORF 2	1301	substitution	C	no
cp000052	ORF 2	1607	substitution	G	no
CTA	ORF 2	1607	substitution	G	no
CTB	ORF 2	1607	substitution	G	no
JALI	ORF 2	1607	substitution	G	no
NC_001372	ORF 2	1607	substitution	A	no
SW2	ORF 2	1607	substitution	A	no
SW3	ORF 2	1607	substitution	A	no
X06707	ORF 2	1607	substitution	A	no
X07547	ORF 2	1607	substitution	A	no
L2 LON	ORF 2	1607	substitution	A	no
cp000052	ORF 2	1637	substitution	T	no
CTA	ORF 2	1637	substitution	T	no
CTB	ORF 2	1637	substitution	T	no
JALI	ORF 2	1637	substitution	T	no
NC_001372	ORF 2	1637	substitution	T	no
SW2	ORF 2	1637	substitution	T	no
SW3	ORF 2	1637	substitution	T	no
X06707	ORF 2	1637	substitution	C	no
X07547	ORF 2	1637	substitution	C	no
L2 LON	ORF 2	1637	substitution	C	no

cp000052	ORF 2	1664	substitution	C	no
CTA	ORF 2	1664	substitution	C	no
CTB	ORF 2	1664	substitution	C	no
JALI	ORF 2	1664	substitution	C	no
NC_001372	ORF 2	1664	substitution	C	no
SW2	ORF 2	1664	substitution	C	no
SW3	ORF 2	1664	substitution	C	no
X06707	ORF 2	1664	substitution	T	no
X07547	ORF 2	1664	substitution	T	no
L2 LON	ORF 2	1664	substitution	T	no
cp000052	ORF 2	1856	substitution	G	no
CTA	ORF 2	1856	substitution	G	no
CTB	ORF 2	1856	substitution	G	no
JALI	ORF 2	1856	substitution	G	no
NC_001372	ORF 2	1856	substitution	G	no
SW2	ORF 2	1856	substitution	G	no
SW3	ORF 2	1856	substitution	G	no
X06707	ORF 2	1856	substitution	A	no
X07547	ORF 2	1856	substitution	A	no
L2 LON	ORF 2	1856	substitution	A	no
cp000052	ORF 2	1904	substitution	A	no
CTA	ORF 2	1904	substitution	A	no
CTB	ORF 2	1904	substitution	A	no
JALI	ORF 2	1904	substitution	A	no
NC_001372	ORF 2	1904	substitution	G	no
SW2	ORF 2	1904	substitution	G	no
SW3	ORF 2	1904	substitution	G	no
X06707	ORF 2	1904	substitution	G	no
X07547	ORF 2	1904	substitution	G	no
L2 LON	ORF 2	1904	substitution	G	no
cp000052	ORF 2	2075	substitution	G	no
CTA	ORF 2	2075	substitution	G	no
CTB	ORF 2	2075	substitution	G	no
JALI	ORF 2	2075	substitution	G	no
NC_001372	ORF 2	2075	substitution	A	no
SW2	ORF 2	2075	substitution	A	no
SW3	ORF 2	2075	substitution	A	no
X06707	ORF 2	2075	substitution	G	no
X07547	ORF 2	2075	substitution	G	no
L2 LON	ORF 2	2075	substitution	G	no

cp000052	INT	2153	substitution	C	no
CTA	INT	2153	substitution	C	no
CTB	INT	2153	substitution	C	no
JALI	INT	2153	substitution	C	no
NC_001372	INT	2153	substitution	T	no
SW2	INT	2153	substitution	T	no
SW3	INT	2153	substitution	T	no
X06707	INT	2153	substitution	T	no
X07547	INT	2153	substitution	T	no
L2 LON	INT	2153	substitution	T	no

cp000052	INT				
CTA	INT				
CTB	INT				
JALI	INT				
NC_001372	INT				
SW2	INT	2225	insertion 44bp		no
SW3	INT				
X06707	INT				
X07547	INT				
L2 LON	INT				

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF3	2355	substitution	C	no
CTA	ORF3	2355	substitution	C	no
CTB	ORF3	2355	substitution	C	no
JALI	ORF3	2355	substitution	C	no
NC_001372	ORF3	2355	substitution	T	no
SW2	ORF3	2355	substitution	T	no
SW3	ORF3	2355	substitution	T	no
X06707	ORF3	2355	substitution	T	no
X07547	ORF3	2355	substitution	T	no
L2 LON	ORF3	2355	substitution	T	no
cp000052	ORF3	2386	substitution	T	tyr
CTA	ORF3	2386	substitution	T	tyr
CTB	ORF3	2386	substitution	T	tyr
JALI	ORF3	2386	substitution	T	tyr
NC_001372	ORF3	2386	substitution	C	his
SW2	ORF3	2386	substitution	T	tyr
SW3	ORF3	2386	substitution	T	tyr
X06707	ORF3	2386	substitution	T	tyr
X07547	ORF3	2386	substitution	T	tyr
L2 LON	ORF3	2386	substitution	T	tyr
cp000052	ORF3	2583	substitution	C	no
CTA	ORF3	2583	substitution	C	no
CTB	ORF3	2583	substitution	C	no
JALI	ORF3	2583	substitution	C	no
NC_001372	ORF3	2583	substitution	T	no
SW2	ORF3	2583	substitution	T	no
SW3	ORF3	2583	substitution	T	no
X06707	ORF3	2583	substitution	T	no
X07547	ORF3	2583	substitution	T	no
L2 LON	ORF3	2583	substitution	T	no
cp000052	ORF3	2676	substitution	T	no
CTA	ORF3	2676	substitution	T	no
CTB	ORF3	2676	substitution	T	no
JALI	ORF3	2676	substitution	T	no
NC_001372	ORF3	2676	substitution	T	no
SW2	ORF3	2676	substitution	T	no
SW3	ORF3	2676	substitution	T	no
X06707	ORF3	2676	substitution	C	no
X07547	ORF3	2676	substitution	C	no
L2 LON	ORF3	2676	substitution	C	no

cp000052	ORF3	2786	substitution	C	ala
CTA	ORF3	2786	substitution	C	ala
CTB	ORF3	2786	substitution	C	ala
JALI	ORF3	2786	substitution	C	ala
NC_001372	ORF3	2786	substitution	T	val
SW2	ORF3	2786	substitution	T	val
SW3	ORF3	2786	substitution	T	val
X06707	ORF3	2786	substitution	C	ala
X07547	ORF3	2786	substitution	C	ala
L2 LON	ORF3	2786	substitution	C	ala
cp000052	ORF3	2788	substitution	G	ala
CTA	ORF3	2788	substitution	G	ala
CTB	ORF3	2788	substitution	G	ala
JALI	ORF3	2788	substitution	G	ala
NC_001372	ORF3	2788	substitution	G	ala
SW2	ORF3	2788	substitution	G	ala
SW3	ORF3	2788	substitution	G	ala
X06707	ORF3	2788	substitution	A	thr
X07547	ORF3	2788	substitution	A	thr
L2 LON	ORF3	2788	substitution	A	thr
cp000052	ORF3	2822	substitution	G	gly
CTA	ORF3	2822	substitution	G	gly
CTB	ORF3	2822	substitution	G	gly
JALI	ORF3	2822	substitution	G	gly
NC_001372	ORF3	2822	substitution	G	gly
SW2	ORF3	2822	substitution	G	gly
SW3	ORF3	2822	substitution	G	gly
X06707	ORF3	2822	substitution	C	ala
X07547	ORF3	2822	substitution	C	ala
L2 LON	ORF3	2822	substitution	C	ala
cp000052	ORF3	2832	substitution	A	no
CTA	ORF3	2832	substitution	A	no
CTB	ORF3	2832	substitution	A	no
JALI	ORF3	2832	substitution	G	no
NC_001372	ORF3	2832	substitution	A	no
SW2	ORF3	2832	substitution	A	no
SW3	ORF3	2832	substitution	A	no
X06707	ORF3	2832	substitution	A	no
X07547	ORF3	2832	substitution	A	no
L2 LON	ORF3	2832	substitution	A	no

cp000052	ORF3	2901	substitution	G	no
CTA	ORF3	2901	substitution	G	no
CTB	ORF3	2901	substitution	G	no
JALI	ORF3	2901	substitution	G	no
NC_001372	ORF3	2901	substitution	A	no
SW2	ORF3	2901	substitution	A	no
SW3	ORF3	2901	substitution	A	no
X06707	ORF3	2901	substitution	G	no
X07547	ORF3	2901	substitution	G	no
L2 LON	ORF3	2901	substitution	G	no
cp000052	ORF3	2919	substitution	A	no
CTA	ORF3	2919	substitution	A	no
CTB	ORF3	2919	substitution	A	no
JALI	ORF3	2919	substitution	A	no
NC_001372	ORF3	2919	substitution	A	no
SW2	ORF3	2919	substitution	A	no
SW3	ORF3	2919	substitution	A	no
X06707	ORF3	2919	substitution	T	no
X07547	ORF3	2919	substitution	T	no
L2 LON	ORF3	2919	substitution	T	no
cp000052	ORF3	3114	substitution	A	no
CTA	ORF3	3114	substitution	A	no
CTB	ORF3	3114	substitution	A	no
JALI	ORF3	3114	substitution	A	no
NC_001372	ORF3	3114	substitution	G	no
SW2	ORF3	3114	substitution	G	no
SW3	ORF3	3114	substitution	G	no
X06707	ORF3	3114	substitution	A	no
X07547	ORF3	3114	substitution	A	no
L2 LON	ORF3	3114	substitution	A	no
cp000052	ORF3	3165	substitution	T	no
CTA	ORF3	3165	substitution	T	no
CTB	ORF3	3165	substitution	T	no
JALI	ORF3	3165	substitution	T	no
NC_001372	ORF3	3165	substitution	C	no
SW2	ORF3	3165	substitution	C	no
SW3	ORF3	3165	substitution	C	no
X06707	ORF3	3165	substitution	T	no
X07547	ORF3	3165	substitution	T	no
L2 LON	ORF3	3165	substitution	T	no

cp000052	ORF3	3185	substitution	A	gln
CTA	ORF3	3185	substitution	A	gln
CTB	ORF3	3185	substitution	A	gln
JALI	ORF3	3185	substitution	A	gln
NC_001372	ORF3	3185	substitution	G	arg
SW2	ORF3	3185	substitution	G	arg
SW3	ORF3	3185	substitution	G	arg
X06707	ORF3	3185	substitution	G	arg
X07547	ORF3	3185	substitution	G	arg
L2 LON	ORF3	3185	substitution	G	arg
cp000052	ORF3	3339	substitution	T	no
CTA	ORF3	3339	substitution	T	no
CTB	ORF3	3339	substitution	T	no
JALI	ORF3	3339	substitution	T	no
NC_001372	ORF3	3339	substitution	T	no
SW2	ORF3	3339	substitution	T	no
SW3	ORF3	3339	substitution	T	no
X06707	ORF3	3339	substitution	C	no
X07547	ORF3	3339	substitution	T	no
L2 LON	ORF3	3339	substitution	T	no
cp000052	ORF3	3579	END		
CTA	ORF3	3579	END		
CTB	ORF3	3579	END		
JALI	ORF3	3579	END		
NC_001372	ORF3	3579	END		
SW2	ORF3	3579	END		
SW3	ORF3	3579	END		
X06707	ORF3	3579	END		
X07547	ORF3	3579	END		
L2 LON	ORF3	3579	END		

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF 4	3758	substitution	A	no
CTA	ORF 4	3758	substitution	A	no
CTB	ORF 4	3758	substitution	A	no
JALI	ORF 4	3758	substitution	A	no
NC_001372	ORF 4	3758	substitution	A	no
SW2	ORF 4	3758	substitution	A	no
SW3	ORF 4	3758	substitution	A	no
X06707	ORF 4	3758	substitution	G	no
X07547	ORF 4	3758	substitution	G	no
L2 LON	ORF 4	3758	substitution	G	no
cp000052	ORF 4	3764	substitution	C	no
CTA	ORF 4	3764	substitution	C	no
CTB	ORF 4	3764	substitution	C	no
JALI	ORF 4	3764	substitution	C	no
NC_001372	ORF 4	3764	substitution	C	no
SW2	ORF 4	3764	substitution	C	no
SW3	ORF 4	3764	substitution	C	no
X06707	ORF 4	3764	substitution	T	no
X07547	ORF 4	3764	substitution	T	no
L2 LON	ORF 4	3764	substitution	T	no
cp000052	ORF 4	3798	substitution	A	Ile
CTA	ORF 4	3798	substitution	A	Ile
CTB	ORF 4	3798	substitution	A	Ile
JALI	ORF 4	3798	substitution	A	Ile
NC_001372	ORF 4	3798	substitution	G	val
SW2	ORF 4	3798	substitution	G	val
SW3	ORF 4	3798	substitution	G	val
X06707	ORF 4	3798	substitution	G	val
X07547	ORF 4	3798	substitution	G	val
L2 LON	ORF 4	3798	substitution	G	val
cp000052	ORF 4	4140	substitution	T	ser
CTA	ORF 4	4140	substitution	T	ser
CTB	ORF 4	4140	substitution	T	ser
JALI	ORF 4	4140	substitution	T	ser
NC_001372	ORF 4	4140	substitution	C	pro
SW2	ORF 4	4140	substitution	C	pro
SW3	ORF 4	4140	substitution	C	pro
X06707	ORF 4	4140	substitution	T	ser
X07547	ORF 4	4140	substitution	T	ser
L2 LON	ORF 4	4140	substitution	T	ser

cp000052	ORF 4	4253	substitution	T	no
CTA	ORF 4	4253	substitution	T	no
CTB	ORF 4	4253	substitution	T	no
JALI	ORF 4	4253	substitution	T	no
NC_001372	ORF 4	4253	substitution	T	no
SW2	ORF 4	4253	substitution	T	no
SW3	ORF 4	4253	substitution	T	no
X06707	ORF 4	4253	substitution	C	no
X07547	ORF 4	4253	substitution	C	no
L2 LON	ORF 4	4253	substitution	C	no
cp000052	ORF 4	4364	substitution	A	leu
CTA	ORF 4	4364	substitution	A	leu
CTB	ORF 4	4364	substitution	A	leu
JALI	ORF 4	4364	substitution	A	leu
NC_001372	ORF 4	4364	substitution	A	leu
SW2	ORF 4	4364	substitution	A	leu
SW3	ORF 4	4364	substitution	A	leu
X06707	ORF 4	4364	substitution	C	phe
X07547	ORF 4	4364	substitution	C	phe
L2 LON	ORF 4	4364	substitution	C	phe
cp000052	ORF 4	4418	substitution	C	no
CTA	ORF 4	4418	substitution	C	no
CTB	ORF 4	4418	substitution	C	no
JALI	ORF 4	4418	substitution	C	no
NC_001372	ORF 4	4418	substitution	C	no
SW2	ORF 4	4418	substitution	C	no
SW3	ORF 4	4418	substitution	C	no
X06707	ORF 4	4418	substitution	T	no
X07547	ORF 4	4418	substitution	T	no
L2 LON	ORF 4	4418	substitution	T	no
cp000052	ORF 4	4479	substitution	C	gln
CTA	ORF 4	4479	substitution	C	gln
CTB	ORF 4	4479	substitution	C	gln
JALI	ORF 4	4479	substitution	C	gln
NC_001372	ORF 4	4479	substitution	C	gln
SW2	ORF 4	4479	substitution	A	lys
SW3	ORF 4	4479	substitution	C	gln
X06707	ORF 4	4479	substitution	C	gln
X07547	ORF 4	4479	substitution	C	gln
L2 LON	ORF 4	4479	substitution	C	gln

cp000052	ORF 4	4487	substitution	C	no
CTA	ORF 4	4487	substitution	C	no
CTB	ORF 4	4487	substitution	C	no
JALI	ORF 4	4487	substitution	C	no
NC_001372	ORF 4	4487	substitution	C	no
SW2	ORF 4	4487	substitution	C	no
SW3	ORF 4	4487	substitution	C	no
X06707	ORF 4	4487	substitution	T	no
X07547	ORF 4	4487	substitution	T	no
L2 LON	ORF 4	4487	substitution	T	no
cp000052	ORF 4	4508	substitution	T	no
CTA	ORF 4	4508	substitution	T	no
CTB	ORF 4	4508	substitution	T	no
JALI	ORF 4	4508	substitution	T	no
NC_001372	ORF 4	4508	substitution	C	no
SW2	ORF 4	4508	substitution	C	no
SW3	ORF 4	4508	substitution	C	no
X06707	ORF 4	4508	substitution	C	no
X07547	ORF 4	4508	substitution	C	no
L2 LON	ORF 4	4508	substitution	C	no
cp000052	ORF 4	4571	substitution	C	no
CTA	ORF 4	4571	substitution	C	no
CTB	ORF 4	4571	substitution	C	no
JALI	ORF 4	4571	substitution	C	no
NC_001372	ORF 4	4571	substitution	C	no
SW2	ORF 4	4571	substitution	C	no
SW3	ORF 4	4571	substitution	C	no
X06707	ORF 4	4571	substitution	T	no
X07547	ORF 4	4571	substitution	T	no
L2 LON	ORF 4	4571	substitution	T	no
cp000052	ORF 4	4586	substitution	T	no
CTA	ORF 4	4586	substitution	T	no
CTB	ORF 4	4586	substitution	T	no
JALI	ORF 4	4586	substitution	T	no
NC_001372	ORF 4	4586	substitution	T	no
SW2	ORF 4	4586	substitution	T	no
SW3	ORF 4	4586	substitution	T	no
X06707	ORF 4	4586	substitution	C	no
X07547	ORF 4	4586	substitution	T	no
L2 LON	ORF 4	4586	substitution	T	no

cp000052	ORF 4	4637	END		
CTA	ORF 4	4637	END		
CTB	ORF 4	4637	END		
JALI	ORF 4	4637	END		
NC_001372	ORF 4	4637	END		
SW2	ORF 4	4637	END		
SW3	ORF 4	4637	END		
X06707	ORF 4	4637	END		
X07547	ORF 4	4637	END		
L2 LON	ORF 4	4637	END		
cp000052	INT	4644	substitution	G	no
CTA	INT	4644	substitution	G	no
CTB	INT	4644	substitution	G	no
JALI	INT	4644	substitution	G	no
NC_001372	INT	4644	substitution	G	no
SW2	INT	4644	substitution	G	no
SW3	INT	4644	substitution	G	no
X06707	INT	4644	substitution	A	no
X07547	INT	4644	substitution	A	no
L2 LON	INT	4644	substitution	A	no

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF5	4732	substitution	G	glu
CTA	ORF5	4732	substitution	G	glu
CTB	ORF5	4732	substitution	G	glu
JALI	ORF5	4732	substitution	G	glu
NC_001372	ORF5	4732	substitution	G	glu
SW2	ORF5	4732	substitution	G	glu
SW3	ORF5	4732	substitution	G	glu
X06707	ORF5	4732	substitution	C	gln
X07547	ORF5	4732	substitution	C	gln
L2 LON	ORF5	4732	substitution	C	gln
cp000052	ORF5	4879	substitution	T	ser
CTA	ORF5	4879	substitution	T	ser
CTB	ORF5	4879	substitution	T	ser
JALI	ORF5	4879	substitution	T	ser
NC_001372	ORF5	4879	substitution	T	ser
SW2	ORF5	4879	substitution	T	ser
SW3	ORF5	4879	substitution	T	ser
X06707	ORF5	4879	substitution	C	pro
X07547	ORF5	4879	substitution	C	pro
L2 LON	ORF5	4879	substitution	C	pro
cp000052	ORF5	4967	substitution	A	asp
CTA	ORF5	4967	substitution	A	asp
CTB	ORF5	4967	substitution	A	asp
JALI	ORF5	4967	substitution	A	asp
NC_001372	ORF5	4967	substitution	A	asp
SW2	ORF5	4967	substitution	A	asp
SW3	ORF5	4967	substitution	A	asp
X06707	ORF5	4967	substitution	G	gly
X07547	ORF5	4967	substitution	G	gly
L2 LON	ORF5	4967	substitution	G	gly
cp000052	ORF5	4994	substitution	A	asn
CTA	ORF5	4994	substitution	A	asn
CTB	ORF5	4994	substitution	A	asn
JALI	ORF5	4994	substitution	A	asn
NC_001372	ORF5	4994	substitution	G	ser
SW2	ORF5	4994	substitution	G	ser
SW3	ORF5	4994	substitution	G	ser
X06707	ORF5	4994	substitution	G	ser
X07547	ORF5	4994	substitution	G	ser
L2 LON	ORF5	4994	substitution	G	ser

cp000052	ORF5	5024	substitution	A	lys
CTA	ORF5	5024	substitution	A	lys
CTB	ORF5	5024	substitution	A	lys
JALI	ORF5	5024	substitution	A	lys
NC_001372	ORF5	5024	substitution	A	lys
SW2	ORF5	5024	substitution	A	lys
SW3	ORF5	5024	substitution	A	lys
X06707	ORF5	5024	substitution	C	thr
X07547	ORF5	5024	substitution	C	thr
L2 LON	ORF5	5024	substitution	C	thr
cp000052	ORF5	5112	substitution	T	ser
CTA	ORF5	5112	substitution	T	ser
CTB	ORF5	5112	substitution	T	ser
JALI	ORF5	5112	substitution	T	ser
NC_001372	ORF5	5112	substitution	T	ser
SW2	ORF5	5112	substitution	T	ser
SW3	ORF5	5112	substitution	T	ser
X06707	ORF5	5112	substitution	G	arg
X07547	ORF5	5112	substitution	G	arg
L2 LON	ORF5	5112	substitution	G	arg
cp000052	ORF5	5114	substitution	G	ser
CTA	ORF5	5114	substitution	G	ser
CTB	ORF5	5114	substitution	G	ser
JALI	ORF5	5114	substitution	G	ser
NC_001372	ORF5	5114	substitution	A	asn
SW2	ORF5	5114	substitution	A	asn
SW3	ORF5	5114	substitution	A	asn
X06707	ORF5	5114	substitution	A	asn
X07547	ORF5	5114	substitution	A	asn
L2 LON	ORF5	5114	substitution	A	asn
cp000052	ORF5	5270	substitution	G	ser
CTA	ORF5	5270	substitution	G	ser
CTB	ORF5	5270	substitution	G	ser
JALI	ORF5	5270	substitution	G	ser
NC_001372	ORF5	5270	substitution	G	ser
SW2	ORF5	5270	substitution	G	ser
SW3	ORF5	5270	substitution	G	ser
X06707	ORF5	5270	substitution	A	tyr
X07547	ORF5	5270	substitution	A	tyr
L2 LON	ORF5	5270	substitution	A	tyr

cp000052	ORF5	5299	substitution	G	Ile
CTA	ORF5	5299	substitution	G	Ile
CTB	ORF5	5299	substitution	G	Ile
JALI	ORF5	5299	substitution	G	Ile
NC_001372	ORF5	5299	substitution	A	val
SW2	ORF5	5299	substitution	A	val
SW3	ORF5	5299	substitution	A	val
X06707	ORF5	5299	substitution	G	Ile
X07547	ORF5	5299	substitution	G	Ile
L2 LON	ORF5	5299	substitution	G	Ile
cp000052	ORF5	5328	substitution	C	arg
CTA	ORF5	5328	substitution	C	arg
CTB	ORF5	5328	substitution	C	arg
JALI	ORF5	5328	substitution	C	arg
NC_001372	ORF5	5328	substitution	T	ser
SW2	ORF5	5328	substitution	T	ser
SW3	ORF5	5328	substitution	T	ser
X06707	ORF5	5328	substitution	A	arg
X07547	ORF5	5328	substitution	A	arg
L2 LON	ORF5	5328	substitution	A	arg
cp000052	ORF5	5333	substitution	C	thr
CTA	ORF5	5333	substitution	C	thr
CTB	ORF5	5333	substitution	C	thr
JALI	ORF5	5333	substitution	C	thr
NC_001372	ORF5	5333	substitution	C	thr
SW2	ORF5	5333	substitution	C	thr
SW3	ORF5	5333	substitution	C	thr
X06707	ORF5	5333	substitution	T	Ile
X07547	ORF5	5333	substitution	T	Ile
L2 LON	ORF5	5333	substitution	T	Ile
cp000052	ORF5	5352	substitution	A	no
CTA	ORF5	5352	substitution	A	no
CTB	ORF5	5352	substitution	A	no
JALI	ORF5	5352	substitution	A	no
NC_001372	ORF5	5352	substitution	G	no
SW2	ORF5	5352	substitution	G	no
SW3	ORF5	5352	substitution	G	no
X06707	ORF5	5352	substitution	G	no
X07547	ORF5	5352	substitution	G	no
L2 LON	ORF5	5352	substitution	G	no

cp000052	ORF5	5466	substitution	G	no
CTA	ORF5	5466	substitution	G	no
CTB	ORF5	5466	substitution	G	no
JALI	ORF5	5466	substitution	G	no
NC_001372	ORF5	5466	substitution	A	no
SW2	ORF5	5466	substitution	A	no
SW3	ORF5	5466	substitution	A	no
X06707	ORF5	5466	substitution	G	no
X07547	ORF5	5466	substitution	G	no
L2 LON	ORF5	5466	substitution	G	no
cp000052	ORF5	5469	substitution	A	no
CTA	ORF5	5469	substitution	A	no
CTB	ORF5	5469	substitution	A	no
JALI	ORF5	5469	substitution	A	no
NC_001372	ORF5	5469	substitution	G	no
SW2	ORF5	5469	substitution	G	no
SW3	ORF5	5469	substitution	G	no
X06707	ORF5	5469	substitution	G	no
X07547	ORF5	5469	substitution	G	no
L2 LON	ORF5	5469	substitution	G	no
cp000052	ORF5	5493	END		
CTA	ORF5	5493	END		
CTB	ORF5	5493	END		
JALI	ORF5	5493	END		
NC_001372	ORF5	5493	END		
SW2	ORF5	5493	END		
SW3	ORF5	5493	END		
X06707	ORF5	5493	END		
X07547	ORF5	5493	END		
L2 LON	ORF5	5493	END		
cp000052	INT	5540	substitution	C	no
CTA	INT	5540	substitution	C	no
CTB	INT	5540	substitution	C	no
JALI	INT	5540	substitution	C	no
NC_001372	INT	5540	substitution	C	no
SW2	INT	5540	substitution	C	no
SW3	INT	5540	substitution	C	no
X06707	INT	5540	substitution	A	no
X07547	INT	5540	substitution	A	no
L2 LON	INT	5540	substitution	A	no

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF6	5577	substitution	A	arg
CTA	ORF6	5577	substitution	A	arg
CTB	ORF6	5577	substitution	A	arg
JALI	ORF6	5577	substitution	A	arg
NC_001372	ORF6	5577	substitution	A	arg
SW2	ORF6	5577	substitution	A	arg
SW3	ORF6	5577	substitution	A	arg
X06707	ORF6	5577	substitution	C	ser
X07547	ORF6	5577	substitution	C	ser
L2 LON	ORF6	5577	substitution	C	ser
cp000052	ORF6	5871	END		
CTA	ORF6	5871	END		
CTB	ORF6	5871	END		
JALI	ORF6	5871	END		
NC_001372	ORF6	5871	END		
SW2	ORF6	5871	END		
SW3	ORF6	5871	END		
X06707	ORF6	5871	END		
X07547	ORF6	5871	END		
L2 LON	ORF6	5871	END		
cp000052	INT	5873	substitution	C	no
CTA	INT	5873	substitution	C	no
CTB	INT	5873	substitution	C	no
JALI	INT	5873	substitution	C	no
NC_001372	INT	5873	substitution	C	no
SW2	INT	5873	substitution	C	no
SW3	INT	5873	substitution	C	no
X06707	INT	5873	substitution	T	no
X07547	INT	5873	substitution	T	no
L2 LON	INT	5873	substitution	T	no

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF7	6064	substitution	A	asn
CTA	ORF7	6064	substitution	A	asn
CTB	ORF7	6064	substitution	A	asn
JALI	ORF7	6064	substitution	A	asn
NC_001372	ORF7	6064	substitution	G	his
SW2	ORF7	6064	substitution	G	his
SW3	ORF7	6064	substitution	G	his
X06707	ORF7	6064	substitution	A	asn
X07547	ORF7	6064	substitution	A	asn
L2 LON	ORF7	6064	substitution	A	asn
cp000052	ORF7	6078	substitution	G	phe
CTA	ORF7	6078	substitution	G	phe
CTB	ORF7	6078	substitution	G	phe
JALI	ORF7	6078	substitution	G	phe
NC_001372	ORF7	6078	substitution	G	phe
SW2	ORF7	6078	substitution	G	phe
SW3	ORF7	6078	substitution	G	phe
X06707	ORF7	6078	substitution	A	leu
X07547	ORF7	6078	substitution	A	leu
L2 LON	ORF7	6078	substitution	A	leu
cp000052	ORF7	6167	substitution	T	leu
CTA	ORF7	6167	substitution	T	leu
CTB	ORF7	6167	substitution	T	leu
JALI	ORF7	6167	substitution	T	leu
NC_001372	ORF7	6167	substitution	C	ser
SW2	ORF7	6167	substitution	C	ser
SW3	ORF7	6167	substitution	C	ser
X06707	ORF7	6167	substitution	T	leu
X07547	ORF7	6167	substitution	T	leu
L2 LON	ORF7	6167	substitution	T	leu
cp000052	ORF7	6335	substitution	T	val
CTA	ORF7	6335	substitution	T	val
CTB	ORF7	6335	substitution	T	val
JALI	ORF7	6335	substitution	T	val
NC_001372	ORF7	6335	substitution	C	ala
SW2	ORF7	6335	substitution	C	ala
SW3	ORF7	6335	substitution	C	ala
X06707	ORF7	6335	substitution	C	ala
X07547	ORF7	6335	substitution	C	ala
L2 LON	ORF7	6335	substitution	C	ala

cp000052	ORF7	6692	END
CTA	ORF7	6692	END
CTB	ORF7	6692	END
JALI	ORF7	6692	END
NC_001372	ORF7	6692	END
SW2	ORF7	6692	END
SW3	ORF7	6692	END
X06707	ORF7	6692	END
X07547	ORF7	6692	END
L2 LON	ORF7	6692	END

sequence	orf	position	type of change	bp change	Pro change
cp000052	ORF8	6714	substitution	G	asp
CTA	ORF8	6714	substitution	G	asp
CTB	ORF8	6714	substitution	G	asp
JALI	ORF8	6714	substitution	G	asp
NC_001372	ORF8	6714	substitution	G	asp
SW2	ORF8	6714	substitution	G	asp
SW3	ORF8	6714	substitution	G	asp
X06707	ORF8	6714	substitution	A	asn
X07547	ORF8	6714	substitution	A	asn
L2 LON	ORF8	6714	substitution	A	asn
cp000052	ORF8	6750	substitution	C	no
CTA	ORF8	6750	substitution	C	no
CTB	ORF8	6750	substitution	C	no
JALI	ORF8	6750	substitution	C	no
NC_001372	ORF8	6750	substitution	C	no
SW2	ORF8	6750	substitution	C	no
SW3	ORF8	6750	substitution	C	no
X06707	ORF8	6750	substitution	T	no
X07547	ORF8	6750	substitution	T	no
L2 LON	ORF8	6750	substitution	T	no
cp000052	ORF8	6767	substitution	A	no
CTA	ORF8	6767	substitution	A	no
CTB	ORF8	6767	substitution	A	no
JALI	ORF8	6767	substitution	A	no
NC_001372	ORF8	6767	substitution	G	no
SW2	ORF8	6767	substitution	G	no
SW3	ORF8	6767	substitution	G	no
X06707	ORF8	6767	substitution	G	no
X07547	ORF8	6767	substitution	G	no
L2 LON	ORF8	6767	substitution	G	no
cp000052	ORF8	6773	substitution	T	no
CTA	ORF8	6773	substitution	T	no
CTB	ORF8	6773	substitution	T	no
JALI	ORF8	6773	substitution	T	no
NC_001372	ORF8	6773	substitution	C	no
SW2	ORF8	6773	substitution	C	no
SW3	ORF8	6773	substitution	C	no
X06707	ORF8	6773	substitution	T	no
X07547	ORF8	6773	substitution	T	no
L2 LON	ORF8	6773	substitution	T	no

cp000052	ORF8	6829	substitution	G	no
CTA	ORF8	6829	substitution	G	no
CTB	ORF8	6829	substitution	G	no
JALI	ORF8	6829	substitution	G	no
NC_001372	ORF8	6829	substitution	G	no
SW2	ORF8	6829	substitution	G	no
SW3	ORF8	6829	substitution	G	no
X06707	ORF8	6829	substitution	A	no
X07547	ORF8	6829	substitution	A	no
L2 LON	ORF8	6829	substitution	A	no
cp000052	ORF8	7355	substitution	A	Ile
CTA	ORF8	7355	substitution	A	Ile
CTB	ORF8	7355	substitution	A	Ile
JALI	ORF8	7355	substitution	A	Ile
NC_001372	ORF8	7355	substitution	G	met
SW2	ORF8	7355	substitution	G	met
SW3	ORF8	7355	substitution	G	met
X06707	ORF8	7355	substitution	G	met
X07547	ORF8	7355	substitution	G	met
L2 LON	ORF8	7355	substitution	G	met
cp000052	ORF8	7433	END		
CTA	ORF8	7433	END		
CTB	ORF8	7433	END		
JALI	ORF8	7433	END		
NC_001372	ORF8	7433	END		
SW2	ORF8	7433	END		
SW3	ORF8	7433	END		
X06707	ORF8	7433	END		
X07547	ORF8	7433	END		
L2 LON	ORF8	7433	END		
cp000052	INT				
CTA	INT				
CTB	INT				
JALI	INT				
NC_001372	INT	7457	insertion 2bp	AT	no
SW2	INT	7457	insertion 2bp	AT	no
SW3	INT	7457	insertion 2bp	AT	no
X06707	INT				
X07547	INT				
L2 LON	INT				

cp000052	INT	7458	substitution	T	no
CTA	INT	7458	substitution	T	no
CTB	INT	7458	substitution	T	no
JALI	INT	7458	substitution	T	no
NC_001372	INT	7458	substitution	C	no
SW2	INT	7458	substitution	C	no
SW3	INT	7458	substitution	C	no
X06707	INT	7458	substitution	C	no
X07547	INT	7458	substitution	C	no
L2 LON	INT	7458	substitution	C	no
cp000052	INT	7472	insertion 2bp	TT	no
CTA	INT	7472	insertion 2bp	TT	no
CTB	INT	7472	insertion 2bp	TT	no
JALI	INT	7472	insertion 2bp	TT	no
NC_001372	INT				
SW2	INT				
SW3	INT				
X06707	INT				
X07547	INT				
L2 LON	INT				
cp000052	INT	7489	substitution	T	no
CTA	INT	7489	substitution	T	no
CTB	INT	7489	substitution	T	no
JALI	INT	7489	substitution	T	no
NC_001372	INT	7489	substitution	C	no
SW2	INT	7489	substitution	C	no
SW3	INT	7489	substitution	C	no
X06707	INT	7489	substitution	T	no
X07547	INT	7489	substitution	T	no
L2 LON	INT	7489	substitution	T	no
cp000052	INT				
CTA	INT				
CTB	INT				
JALI	INT	7500	insertion 4bp	GCAC	no
NC_001372	INT				
SW2	INT				
SW3	INT				
X06707	INT				
X07547	INT				
L2 LON	INT				