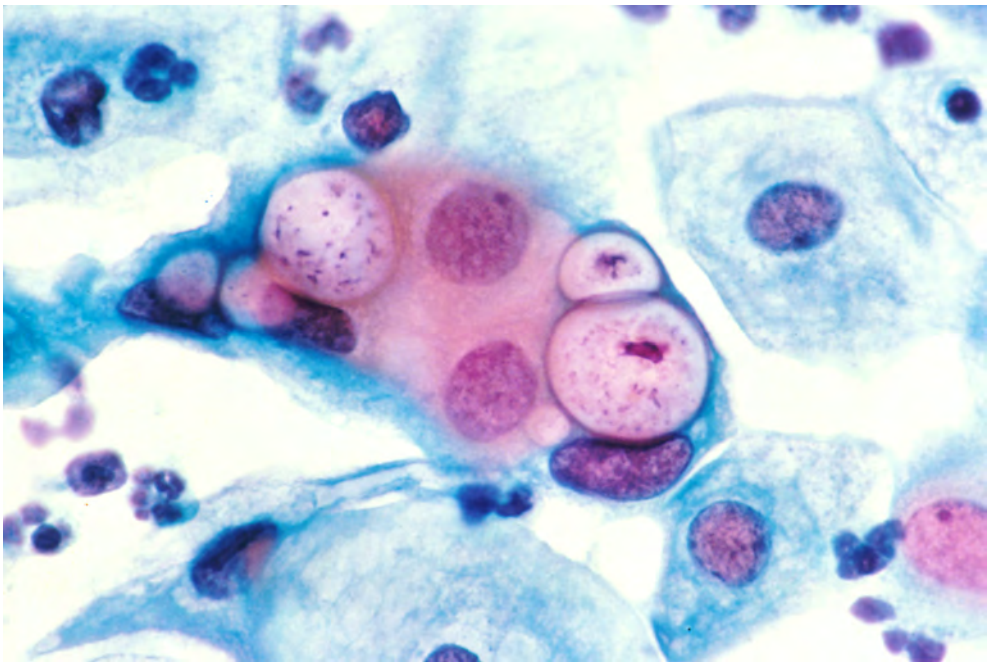
UPPSALA
UNIVERSITET

Genotyping of *Chlamydia trachomatis* strains from patients with different clinical symptoms using a multilocus sequence typing (MLST) system



Linus Christerson

Degree project in biology, Master of science (1 year), 2008
Examensarbete i biologi 30 hp till magisterexamen, 2008
Biology Education Centre, Uppsala University, and Department of Clinical Microbiology, Uppsala University Hospital, 751 85 Uppsala
Supervisors: Björn Herrmann and Markus Klint

Table of contents

1. Summary

The obligate intracellular bacterium *Chlamydia trachomatis* is a major human health concern, causing trachoma and urogential chlamydia infections. Trachoma, which is an eye infection, is endemic in some developing countries and not long ago was the leading cause of preventable blindness, while urogential chlamydia is one of the most common sexually transmitted infections world-wide. A urogenital chlamydia infection often does not cause symptoms, but if left untreated it can lead to serious complications such as pelvic inflammatory disease in women, which can lead to infertility or potentially fatal ectopic pregnancies. In men the infection can spread to the testicles, causing infertility as well. If detected an infection is easily treated with antibiotics. Thus it is of high importance to have adequate tools for reliable diagnostics, molecular epidemiology and contact tracing.

The insufficiency of the existing systems was highlighted when a new genetic variant of *C. trachomatis* emerged in Sweden in 2006. This mutant avoided commonly used detection systems and lead to infected patients being incorrectly reported as uninfected. A novel multilocus sequence typing (MLST) system was created by Klint *et al.* to suit the needs for high resolution molecular epidemiology and contact tracing. This system utilizes five different genetic target regions, which are amplified with PCR, sequenced and subsequently assigned a genetic profile according to a reference database.

The main aim for this project was to examine if a correlation could be established between the genetic profiles generated by the novel MLST system and the clinical symptoms of patients infected with urogenital chlamydia. Such a correlation would have been a way to approach the molecular reason for why complications sometimes occur. Seventy-five Dutch samples were genotyped for this reason, but no correlation could be found, which was not surprising though. The emergence of complications might be due to genetic traits or immunological responses specific to the patients rather than due to the different strains of *C. trachomatis*.

The aims also included expanding the MLST database with more allele variants and characterising and optimising the still novel system. Twenty-four new allele variants were found in total and the MLST system once again showed its unmatched ability to discriminate between different strains of *C. trachomatis*. The system was also optimised by removing redundant allele variants from the database and by shortening one of the target regions.

The MLST system has shown its usefulness in high-resolution molecular epidemiology, but to enable use in routine molecular epidemiology and contact tracing it needs to be simplified and have its throughput increased, perhaps by multiplexing the PCR amplification and by putting the system on a DNA microarray or microbead array.

2. Introduction

Chlamydia trachomatis has a major pathogenic impact on human life, causing trachoma, urogenital chlamydia infections and lymphogranuloma venereum (LGV). Only a decade ago trachoma was the world's leading cause of preventable blindness, but large efforts in containing the endemic have reduced the number of infected individuals to about 84 million (World Health Organization, 2007). Urogenital chlamydia is one of the most common sexually transmitted diseases, with over a million reported infections each year in the United States (Centers for Disease Control and Prevention, 2007) and more than 47 000 reported cases in Sweden each year (Swedish Institute for Infectious Disease Control (Smittskydds-institutet), 2008). If left untreated, urogenital chlamydia can cause pelvic inflammatory disease (PID) in women, which can lead to sterility or potentially fatal ectopic pregnancies. Complications in men are more rare, but the infection can spread to the testicles, causing epididymitis and sterility (Belland *et al*., 2004).

2.1 History

Trachoma was known by the ancient Egyptians as it is described in the Ebers Papyrus dating back to 1550 BC (Taylor, 2008). This papyrus is one of the oldest preserved medical documents and highlights the persistent impact chlamydial infections have had on humans throughout history. In the early 19th century trachoma became a rampant problem in Europe when soldiers from the Napoleonic Wars, infected due to the lack of hygiene in the military camps, returned home and spread the disease to their communities. An improvement in living standards and basic hygiene slowed down the epidemic and by the beginning of the 20th century the disease was more or less under control (Taylor, 2008). In 1938 it was reported that trachoma could be successfully treated with sulphonamide antibiotics (Thygeson, 1938), and by the 1950's trachoma was virtually eliminated from the industrialized world.

The actual disease causing agent was discovered by Halberstaedter and von Prowazek in 1907 when doing trachoma experiments on orangutans (Mabey *et al*., 2003). However, it was long believed to be a virus due to its virus-like characteristics and it was not until 1966 that it was correctly classified as a bacterium (Moulder, 1966). Urogenital chlamydia was soon established as a sexually transmitted disease causing several complications, and during the 1980's and 1990's it became mandatory to report infections in many western countries.

2.2 The *Chlamydiales* and their taxonomy

C. trachomatis belongs to the order *Chlamydiales*. The *Chlamydiales* reside in an exclusive taxonomic position and their unique lifestyle contributes to their unusual pathogenic potential. They are an intriguing group of obligate intracellular bacteria which have a distinctive developmental cycle. They need to be inside the very cells of a host to be able to replicate.

A major revision of the taxonomy of the *Chlamydiales* was proposed in 1999 (Everett *et al*., 1999). This revision was based on sequence similarities of cell surface antigens, and it was later backed up by phylogenetic analysis of the RNase P RNA gene *rnpB* (Herrmann *et al*., 2000) and ribosomal RNA (Bush and Everett, 2001). This stirred much controversy, but the new classification is nevertheless gaining acceptance (Figure 1).

The order *Chlamydiales* have four families: *Chlamydiaceae*, *Parachlamydiaceae*, *Waddliaceae* and *Simkaniaceae*. The last three include species which infect amoebas. *Chlamydiaceae* is further divided into two genera: *Chlamydophila* and *Chlamydia*. *Chlamydophila* include species such as *C. pneumoniae*, which is a common cause of atypical human pneumonia around the world, and *C. psittaci*, which is endemic among birds and can infect humans causing respiratory psittacosis, also known as parrot fever. The genus *Chlamydia* contains three species: *C. trachomatis*, the human pathogen, *C. suis,* which infects swine, and *C. muridarum* which infects mice and hamsters.

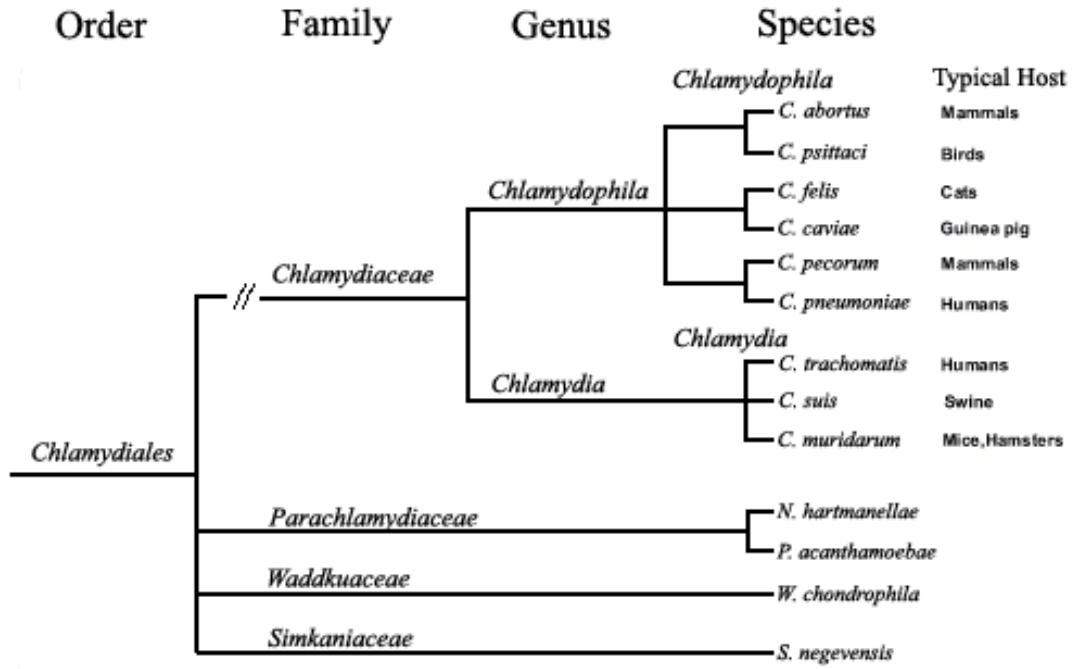


Figure 1. Taxonomy of the Chlamydiales. (Adapted from Bush and Everett, 2001.)

C. trachomatis is further divided into 14 groups called serovars (Table 1), based on traditional serology with antibodies directed against cell surface antigens, of which the major outer membrane protein (MOMP) is the most prominent antigen. There are also several subserotypes which have been used in varying degree in the literature. The serovars can be arranged into two biovars, i.e. two groups with different phenotypical characteristics: the trachoma biovar consisting of serotypes A-K which infect epithelial tissues, causing trachoma (A-C) and urogenital chlamydia infections (D-K), and the LGV biovar consisting of the invasive serotypes L1-3 which infect lymphatic tissues.

Table 1. The 14 main serovars of *C. trachomatis*.

Serovars	Biovars	Disease
A, B, C	Trachoma biovar	Trachoma
D, E, F, G, H, I, J, K	Trachoma biovar	Urogenital chlamydia
L1, L2, L3	LGV biovar	LGV

2.3 Biphasic life cycle of the *Chlamydiales*

All *Chlamydiales* share a unique biphasic life cycle during which the bacterium alters between two distinctly different cell forms: the elementary bodies (EBs) and the reticulate

bodies (RBs). EBs are rigid, metabolically inert, extracellular and infectious. Their nucleoid is tightly packed and their diameter is only 0.2-0.6 μm (Mpiga and Ravaoarinoro, 2006). RBs on the other hand are intracellular. They are more fragile, metabolically active and able to divide by binary fission. Their nucleoid is less compacted and they are larger with a diameter of 0.6-1.5 μm (Mpiga and Ravaoarinoro, 2006).

The cycle begins when EBs attach to and stimulate uptake by a suitable host cell (Figure 2), a process in which two tyrosine kinases, platelet-derived growth factor receptor beta and abelson kinase, recently have been discovered to play an important role (Elwell *et al*., 2008). The EBs remain internalized in vacuoles called inclusions, which are separated from the endocytic pathway, thereby avoiding being fused with phago- or lysosomes (Dautry-Varsat *et al*., 2004). The EBs differentiate into RBs and begin dividing. If the conditions are favourable they will continue dividing and then asynchronously redifferentiate back into EBs, which take about 18 hours in *C. trachomatis*. About 48-72 hours post infection the cell goes through apoptosis, expelling the EBs into the surroundings and initiating a new cycle. In some cases the EBs can be released through exocytosis instead (Mpiga and Ravaoarinoro, 2006). If the host cell is under environmental stress and exposed to interferon-γ the RBs might go into a dormant state as a non-replicating form called persistent bodies (PBs). This can cause a chronic inflammation and lead to serious consequences for the host (Morrison, 2003).

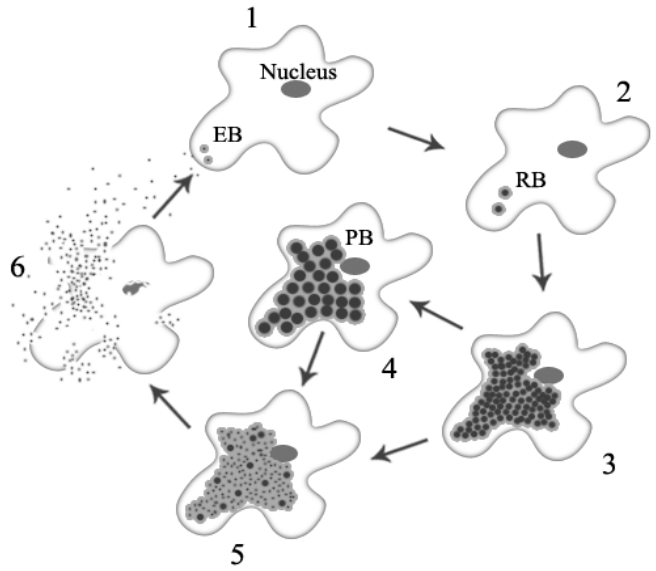


Figure 2. Schematic view of the *Chlamydial* life cycle. (Adapted from Byrne and Ojcius, 2004.) Infectious elementary bodies (EBs) adhere to the cell wall of the host and stimulate uptake into vesicles which avoid phago-lysosomal fusion (1). EBs are transformed into reticulate bodies (RBs) (2) and begin replicating inside their inclusions (3). Under favourable conditions the RBs will redifferentiate back into EBs (5) and lysis of the cell will occur (6), beginning a new round of infection. However, during environmental stress the RBs might change into non-replicating persistent bodies (PBs) (4). When the stress disappears they can redifferentiate into EBs and continue the cycle.

2.4 Urogenital Chlamydia

Urogenital chlamydia is one of the most common sexually transmitted diseases and it is increasing in incidence world-wide with few exceptions. It has been called “the silent epidemic” since it can linger around for years and not cause symptoms until a complication occurs, and then treatment might be too late to stop permanent damage.

2.4.1 Symptoms

Symptoms are somewhat rare, occurring in about 50 % of infected men and in 30 % of infected women (Hafner and McNeilly, 2008). The initial symptoms include a mucous discharge from the penis or vagina and a burning sensation when urinating. As the infection becomes deeper, the symptoms in women extend to a lower abdominal pain, unusual pain during intercourse and bleeding between menstrual periods. In men a deeper infection might show symptoms as pain and swelling of the testicles.

2.4.2 Complications

Complications from an untreated chlamydia infection are more common and severe in women than in men and include pelvic inflammatory disease (PID), which is a term for infections of the uterus, Fallopian tubes and ovaries. It has previously been reported that as many as 40 % of women with untreated chlamydia infections develop PID, but more recent studies have concluded that the incidence is much lower (Low *et al*., 2006). PID causes scarring and fibrosis of the affected tissues. This can block or interrupt the normal movement of the egg and lead to ectopic pregnancies, which is when a fertilized egg starts to grow in the Fallopian tube, which might burst and cause internal bleeding and death. Complete obstruction of the Fallopian tubes prevents sperms from reaching the egg, causing infertility. PID can also lead to chronic pelvic pain. Complications in men are more rare, but the infection can spread to the testicles and cause epididymitis that, if left untreated, can lead to sterility.

Other risks with untreated chlamydia infections include passing the infection from a mother to her baby during childbirth causing an eye infection and/or pneumonia in the newborn. There is also a higher risk of contracting other sexually transmitted diseases, including HIV, if exposed while having an untreated chlamydia infection (Da Ros and Schmitt, 2008).

In rare cases a chlamydia infection can give rise to Reiter's syndrome. This complication is more common among men than women and is a form of arthritis, an inflammation of the joints, and it also involves inflammation of the eyes and mucous membranes, especially in the urogenital tract (Wu and Schwartz, 2008). In some of these patients the arthritis become chronic.

2.4.3 Lymphogranuloma venereum

Lymphogranuloma venereum (LGV) is a more severe sexually transmitted disease caused by the invasive serotypes L1-3. It is endemic in parts of Africa, Latin America, and Asia but only occurred sporadically in Europe and North America prior to a 2003 outbreak in the Netherlands, after which the frequency has been increasing (Klint *et al.*, 2006). In the western world LGV so far has been found almost exclusively among men who have sex with men.

LGV is an infection of the lymphatic system. The symptoms include formation of buboes at the lymph nodes in the groin, or anal bleeding and proctitis, which is an inflammation of the rectal mucosa. Untreated LGV can lead to genital elephantiasis because of lymphatic obstruction, or rectal fistulas, stricture and fibrosis, which might obstruct the bowels and cause them to burst, leading to death.

2.4.4 Treatment

When they have been detected, *C. trachomatis* infections are easily treated with oral antibiotics such as azithromycin, doxycycline or tetracycline. Possible complications are treated individually by other means. *C. trachomatis* infections in Sweden have to be reported to the Swedish Institute for Infectious Disease Control (Smittskyddsinstitutet) and contact tracing is an important part of both treatment and prevention. Re-infection by untreated sex partners seriously increases the risk for complications.

2.5 New variant of *C. trachomatis*

Nucleic amplification tests have been used in clinical diagnosis of *C. trachomatis* since the beginning of the 1990's. Two of the three formerly used tests used in Sweden, the Abott m2000 and the COBAS TaqMan/Amplicore48, were based on a target region in a cryptic plasmid, i.e. a plasmid without known functions.

In the county of Halland the number of reported chlamydia infections suddenly decreased by 25 % between November 2005 and August 2006 after years of constant increase (Ripa and Nilsson, 2006). However, in other parts of Sweden the increase continued. This lead to an investigation where a 377 base pair deletion was detected at the target site in the plasmid, causing two of the three clinically used tests to be unable to amplify the region and thereby give false negatives (Ripa and Nilsson, 2007).

This mutation, causing the strain to remain undetected and untreated, offered the organism an evolutionary advantage compared to non-mutated strains and it was soon discovered that it had spread quickly in Sweden (Soderblom *et al*,. 2006). After re-testing old samples in counties using the Abott or COBAS test it was found that the new variant *C. trachomatis* (nvCT) accounted for 20 - 65 % of all detected chlamydia infections (Herrmann, 2007). About 8000 people were tested negative in total, even though they were infected (Herrmann, 2007). How many additional cases of complications this has caused will never be known. The nvCT is still exclusive to Sweden. Only a few isolated cases have been detected abroad even though there have been a number of large investigations.

2.6 Subtyping of *C. trachomatis*

The traditional way of subtyping *C. trachomatis* is by using a serological approach where antibodies directed against cell surface antigens are used to discriminate different strains. However, this approach depend upon multiple passages of *C. trachomatis* in eukaryotic cell culture and a large panel of antibodies, and is therefore laborious and time-consuming.

The advent of PCR-based amplification meant these problems could be avoided. Serotyping of *C. trachomatis* mainly targets the major outer membrane protein (MOMP), encoded by the gene *ompA*. This is a highly variable gene and most previous genotyping of *C. trachomatis* have been based on *ompA* amplification and subsequent restriction fragment length polymorphism (RFLP) analysis or DNA sequencing.

Phylogenetic studies of *ompA* have shown however that this gene differs in phylogeny and rate of evolution from other regions of the genome, hypothetically due to recombination

events between different strains, and that its phylogenetic characterisation results in phylogenetic groups which show no correlation with biological features such as tissue tropism and disease presentation (Brunelle and Sensabaugh, 2006, Gomes *et al.*, 2007).

In addition, *ompA* genotyping offers only a limited resolution which is unsatisfactory for high resolution molecular epidemiology. Almost half of all urogenital chlamydia infections in Sweden is of serotype E, and within this serotype one specific *ompA* genotype appears to be predominant (Jurstrand *et al*., 2001, Lysén *et al*., 2004). The nvCT belongs to serotype E, and the possibility to track its spread has therefore been restricted. This has highlighted the inadequate resolution of *ompA* genotyping even more.

2.7 The MLST system of *C. trachomatis*

The first multilocus sequence typing (MLST) system was proposed in 1998 using *Neisseria meningitidis* as an example (Maiden, 2006). The method has since then proven to be highly useful and at present there are more than 50 different MLST systems published at http://pubmlst.org/databases.shtml. MLST systems use PCR-amplification of several carefully selected genetic regions, subsequent DNA sequencing and assignment of allelic numbers according to a reference database, which in the end provides the sample with a genetic profile.

The construction of a MLST system for *C. trachomatis* was carried out mainly at the Department of Clinical Microbiology at the Uppsala University Hospital. Through a computational approach the highly conserved genome of *C. trachomatis* was analysed for variable regions. Eight candidate regions were initially identified and their variation was further evaluated in 16 reference strains. After this five of the eight regions were chosen to be included in the MLST system (Klint *et al*., 2007). These five regions consist of three hypothetical genes, i.e. open reading frames: CT058, CT144, CT172, and two known genes, *hctB* and *pbpB* (Figure 3). The regions are named after the gene dominating them.

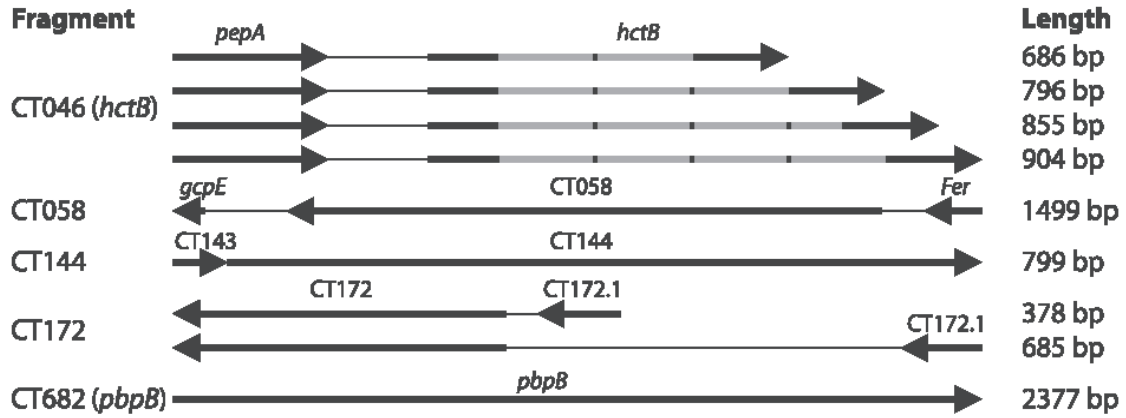


Figure 3. A schematic view of the five target regions finally chosen to be included in the novel MLST system of *C. trachomatis* as defined by the primers used. (Klint *et al.*, 2007.) Bold arrows represent open reading frames. The *hctB* region exists in several length variants. The CT172 region exists in two length variants.

The variation in these regions mainly consists of point mutations. The CT172 region however comprises two length variants and furthermore contains a repetitive number of guanine residues which provides most of its discriminatory capacity. The *hctB* region is unique, with

two to four larger repetitive elements of varying sizes, accounting for the different length variants of the region. These elements in turn are built up by smaller highly similar repetitive segments.

The *hctB* gene encodes a histone H1-like protein that functions as a global regulator of chromatin structure and gene expression. It is important in the transition from RBs to EBs, where the latter one is characterized by densely condensed chromatin (Grieshaber *et al.*, 2006). The *pbpB* gene encodes a penicillin binding protein. The high variability in this gene might be due to a positive selection for the purpose of avoiding the host immune response (Klint *et al*., 2007).

This novel MLST system has been found to be significantly more discriminatory than ordinary *ompA* genotyping. In a collection of 47 clinical samples the MLST system was able to find 32 different variants, while *ompA* analysis only detected 12 variants (Klint *et al.*, 2007).

2.8 Aims

This project had three aims. The main aim was to investigate if a correlation could be established between the genotype profiles generated by the novel MLST system of *C. trachomatis* and the clinical symptoms of patients infected with urogenital chlamydia. A correlation would be a way to approach the molecular reason for why severe complications sometimes occur. It would also be useful if after diagnosis one knew if the strain in question often causes serious complications. Then it would be of high priority to do a more thorough contact tracing. The second aim was to expand the MLST-database with more allele variants, and the third aim was to characterize and optimise the MLST system in order to simplify future use in molecular epidemiology and contact tracing.

3. Results

3.1 Investigation of a hypothetical correlation between MLST genotypes and clinical symptoms

To explore whether a correlation could be established between multilocus sequence typing (MLST) genotypes and clinical symptoms of patients with urogenital chlamydia, samples from 75 Dutch patients with different symptoms were analysed. It was not known in advance which samples were linked to which symptoms, ensuring an objective analysis.

All samples were easily amplified with PCR (Figure 4).

Figure 4. Agarose gel electrophoresis of amplicons from *pbpB*-2 in the Dutch samples. The PCR products for 16 samples, N43-N81, are shown from left to right. The last two wells show water as a negative control and a previously successfully amplified sample as a positive control.

A complete MLST profile was achieved on all 75 samples, as seen in the appendix in Table 8. An example of five samples are displayed in Table 2, showing the allelic numbers assigned to each region after sequencing and consulting the MLST database, as well as results from conventional *ompA*-genotyping as a comparison.

Table 2. An example of five genotyped samples.

Sample ID	CT058	CT144	CT172	*hctB*	*pbpB*	*ompA*	SE genotype
N21	2	6	14	1	2	6	E
N22	2	6	2	35	2	6	E
N23	19	6	2	5	41	6	E
N24	19	7*	2	1	1	6	E
N26	2	6	2	35	2	6	E

The numbers correspond to allele variants in the MLST database. The different MLST profiles are color coded to make it easier to see which ones are identical. White profiles are unique. An asterisk (*) indicates sequences of somewhat lower quality. This has been assessed to have no impact on the results and conclusions (Discussion section 4.1). SE genotype is the genotype attained from *ompA*-genotyping.

Four of the 75 samples were confirmed with real-time PCR to contain *C. trachomatis* strains from both a non-LGV causing serovar as well as the LGV serovar L2. Probes specific for non-LGV and LGV causing serovars respectively both gave signals, showing that there were more than one strain in these samples. One person is also believed to have been double infected with two different non-LGV serovars, as noted in the appendix in Table 18. These five samples were therefore omitted from further analysis.

The different groups of symptoms were asymptomatic, symptomatic and lower abdominal pain (LAP), which indicates a deeper infection and possible complications. The complete results are shown in Table 3, grouped according to the different kinds of symptoms.

Table 3. Complete color coded MLST profiles of samples from patients with clinical symptoms.

Sympt-oms	Sample ID	CT058	CT144	CT172	*hctB*	*pbpB*	ompA	SE genotype	NL genotype
asymp	**N4**	12*	7	2	12	10	1	D	D
asymp	**N5**	26	7	2	5	40	31	D	D
asymp	**N6**	19	7	2	5	37	31	D	D
asymp	**N11**	4	1	4	10	23	32	D	D-
asymp	**N12**	4	1	4	34	23	2	D	D-
asymp	**N13**	4*	1	3	10	23	2	D	D-
asymp	**N19**	28	7	2	5	37	33	D	Da
asymp	**N21**	2	6	14	1	2	6	E	E
asymp	**N22**	2	6	2	35	2	6	E	E
asymp	**N23**	19	6	2	5	41	6	E	E
asymp	**N30**	19	7	1	5	4	24	F	F
asymp	**N31**	19	7	1	5	4	24	F	F
asymp	**N32**	19	7	1	5	4	24	F	F
asymp	**N39**	6	10	1	10	6	9	G	G
asymp	**N40**	6	10	1	10	6	9	G	G
asymp	**N41**	6	10	1	10	6	9	G	G
asymp	**N43**	8	1	4	10	42	11	G	Ga
asymp	**N44**	5	12	3	10	5	11	G	Ga
asymp	**N45**	8	1	3	10	5	11	G	Ga
asymp	**N49**	8	1	4	10	21	18	H	H
asymp	**N50**	29	11	9	12	21	35	H	H
asymp	**N51**	12	11	9	12	21	35	H	H
asymp	**N58**	5	12	7	10	45	36	I	Ia
asymp	**N59**	5	12	7	38	45	36	I	Ia
asymp	**N67**	6	1	7	8	5	37	I	I-
asymp	**N69**	5	12	7	10	45	36	I	I-
asymp	**N76**	12	11	9	12	22	20	J	J
asymp	**N81**	8	9	4	10	8	12	K	K
asymp	**N82**	4	1	7	10	44	12	K	K
asymp	**N83**	7	20	4	33	8	12	K	K
symp	**N1**	8	1	7	8	18	30	B	B
symp	**N7**	27	18	2	7	34	1	D	D
symp	**N8**	19	7	2	5	37	31	D	D
symp	**N9**	19	7	1	5	37	1	D	D
symp	**N14**	4*	9	3	10	23	2	D	D-
symp	**N15**	8	1	4	10	23	2	D	D-
symp	**N20**	8	7	2	5	4	34	D	Da
symp	**N24**	19	7*	2	1	1	6	E	E
symp	**N26**	2	6	2	35	2	6	E	E
symp	**N35**	19	7	1	5	4	24	F	F
symp	**N42**	19	1	1	10	6	11	G	G
symp	**N46**	6	10	1	10	6	9	G	Ga
symp	**N47**	8	1	3	10	5	11	G	Ga
symp	**N53**	12	11	9	12	21	35	H	H
symp	**N54**	29	11	9	12	21	35	H	H
symp	**N61**	5	12	7	10	45	36	I	Ia

symp	**N63**	5	12	7	38	45	36	I	Ia
symp	**N70**	5	12	7	10	45	36	I	I-
symp	**N71**	5	12	7	10	45	36	I	I-
symp	**N78**	4	1	1	10	22	20	J	J
symp	**N84**	7	9	4	10	8	12	K	K
symp	**N85**	7	9	3	10	8	12	K	K
symp	**N86**	7	9	1	10	8	12	K	K
LAP	**N2**	8	1	7	8	18	30	B	B
LAP	**N10**	4	7	1	5	4	1	D	D
LAP	**N16**	8	1	4	10	23	2	D	D-
LAP	**N17**	8	1	4	10	21	2	D	D-
LAP	**N27**	19	7	3	5	1	6	E	E
LAP	**N28**	2	6	14	1	2	6	E	E
LAP	**N29**	19	7	2	7	1	6	E	E
LAP	**N36**	19	1	1	5	4	24	F	F
LAP	**N37**	19	7	1	5	4	24	F	F
LAP	**N38**	19	5	2	5	4	24	F	F
LAP	**N48**	8	1	3	10	5	11	G	Ga
LAP	**N55**	15	7	1	37	43	38	J	H
LAP	**N56**	12	11	9	12	21	35	H	H
LAP	**N57**	15	7	1	36	43	38	J	H
LAP	**N66**	30	1	7	10	5	37	I	Ia
LAP	**N72**	6	1	7	8	5	37	I	I-
LAP	**N87**	12	21	9	12	24	12	K	K

The numbers correspond to allele variants in the MLST database. The different MLST profiles are color coded to facilitate finding which ones are identical. White profiles are unique. An asterisk (*) indicates sequences of somewhat lower quality. This has been assessed to have no impact on the results and conclusions (Discussion section 4.1). SE genotype is the genotype attained from *ompA*-genotyping, while NL genotype is the genotype attained from restriction fragment length polymorphism analysis preformed by Servaas Morre *et al.*, (unpublished data). Strains belonging to the B serovar usually cause trachoma, but can also infect the urogenital tract.

Some MLST profiles differ in only one single point mutation in one region. Therefore a phylogenetic study was carried out, in order to get a better view of how genetically alike the different samples and their corresponding MLST profiles were. A phylogenetic tree was created using the clustering method neighbor-joining. Other methods were also tried and showed similar but not satisfactory results. The parsimony method for example ignores mutations which are unique to a single sample. A schematic phylogram showing the relative genetic distance of the different samples is shown in Figure 5. A boostrapped cladogram showing the reliability of the branching order is shown in Figure 6 in the appendix.

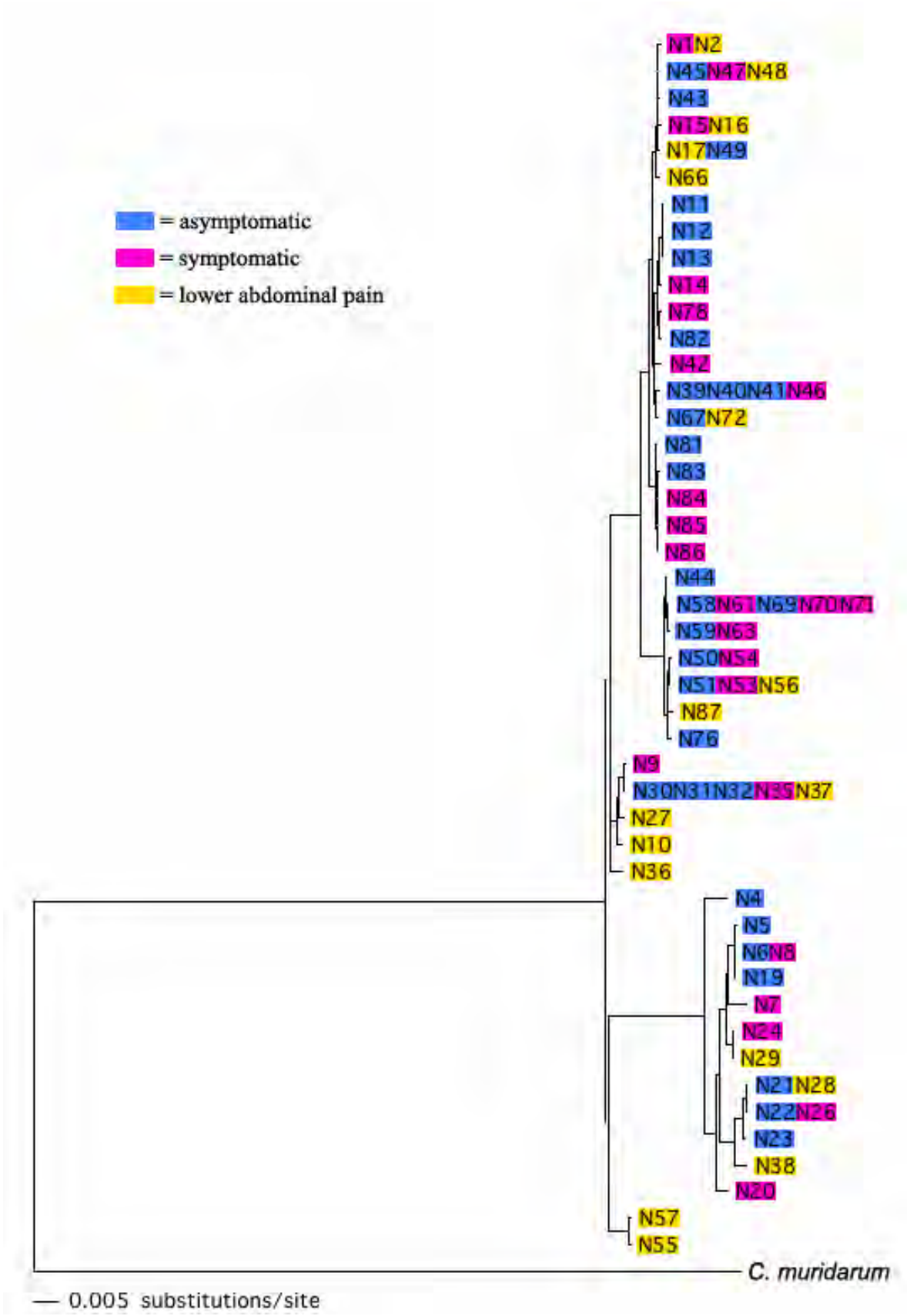


Figure 5. Schematic phylogram based on the five regions included in the MLST system. The tree displays the genetic distance between the Dutch samples. Blue, pink and yellow represent samples from asymptomatic and symptomatic patients and from patients with lower abdominal pain, respectively. The phylogram has been rooted with a corresponding sequence from *C. muridarum* as an outgroup.

3.2 Expanding the MLST database with more allele variants

Previously the novel MLST system of *C. trachomatis* had been applied mostly on Swedish samples, and therefore it was expected to find new variants when sequencing samples from abroad. In total 23 new allele variants were discovered in the Dutch samples (Table 4). Most of the new variants were found in the *hctB* and *pbpB* regions. The new variation was almost exclusively single point mutations, apart from two variants which were more different, as well as two deletion mutations in the *hctB* region. All of the 23 samples were re-amplified and re-sequenced to confirm that the mutations were not PCR artefacts.

Table 4. Overview of new allele variants found in 75 samples from the Netherlands

	CT058	CT144	CT172	*hctB*	*pbpB*
Number of new variants	**5**	**4**	**1**	**6**	**7**

The 75 Dutch samples generated 47 different MLST profiles, out of which 38 were new profiles, i.e. previously not in the database. This can be compared to the traditional *ompA* genotyping which found 18 different genotypes. Complementary additions were also made to another earlier study in which the nvCT was investigated in samples from Örebro, Sweden, colleted in 2006. One more *pbpB* variant was confirmed, consisting of a single point mutation. Thirty samples from France, collected from men who have sex with men and including both non-LGV and LGV serovars, were also investigated in the search for new allele variants. So far, 120 out of 150 alleles have been genotyped, but no new allele variants have been found.

3.3 Optimizing the MLST system

3.3.1 Removal of redundant allele variants in the *hctB* region database

Commonly, sequences from the *hctB* region, like those from other regions, have been automatically aligned using BioEdit 7.0. But due to the repetitive nature of the *htcB* sequences the program cannot carry out a correct alignment. For example, identical sequences of somewhat different length have been completely differently aligned, causing much confusion. The allele variants in the *hctB* database were therefore manually aligned, Figure 7 in the appendix, revealing that four out of the 39 *hctB* allele variants were redundant, i.e. they already existed in the database. The allelic numbers given to these redundant variants will not be reassigned to new variants since that would be misleading with respect to older results where samples were incorrectly genotyped. The allelic numbers assigned to the redundant variants will remain in the system, but simply will not be used anymore.

3.3.2 Shortening of the *pbpB* target region

The *pbp*B region covered 3100 base pairs when the MLST system was first constructed. Due to low success in the PCR amplification of such a large segment, it was later reduced to 2366 base pairs. However, there were still problems, and another reduction was tried where the segment was divided into two smaller separate amplicons spanning 823 and 912 nucleotides respectively, which is what was used temporarily in this project. None of these reductions had been finalized however, and the database contained allele variants of different sizes (Table 5).

This caused problems, because sometimes the variation which separated two allele variants was located outside of the currently amplified region, making the genotyping indecisive.

Table 5. The number of length variants in the *pbpB* database before shortening.

	Length		
	3100 bp	**2366 bp**	**823 and 912 bp**
Number of variants	**31**	**7**	**6**

Thus, a decision was needed: What was most desired? A high success rate but somewhat lower resolution, or high resolution but fewer successfully genotyped samples? In order to take a well-balanced decision, 24 Dutch samples were re-amplified as 2366 base pair fragments to see if any variation had been missed when they were amplified as two smaller separate fragments. One new allele variant was found, which is already counted in Table 4. The total number of variants which would be lost upon shortening the *pbpB* region is summarized in Table 6. In total six out of 44 allele variants would be lost if the *pbpB* region was reduced from 3100 base pair to two separate amplicons of 823 and 912 base pairs respectively.

Table 6. Number of allele variants lost due to a shortening of the *pbpB* region.

	3100 → 2366 bp	**2366 → 823 and 912 bp**	**3000 → 823 and 912 bp**
Number of variants lost	**2**	**4**	**6**

4. Discussion

4.1. No correlation between MLST genotypes and clinical symptoms

There was no correlation between the multilocus sequence typing (MLST) genotypes and the clinical symptoms. It is unnecessary to use more sophisticated statistical methods since none of the MLST profiles containing more then one sample was unique to a specific symptom category. Some profiles even contained samples from all three categories. However, it was not surprising that no correlation could be found. The severe complications of some patients with urogenital chlamydia might be associated with other genes in the bacterial genome, or there might be complex reasons with different factors involved. The complications might even be entirely patient specific and have nothing to do with different strains of *C. trachomatis.*

As noted in the results a few sequences were of somewhat lower quality (marked with an asterisk in the tables). Usually this means that the sequence did not cover the entire target region, e.g. because one of the primers did not work. There was no ambiguity in the identity of the sequences with respect to the existing allele database, but since the entire target region was not covered, there is a theoretical possibility that some new unique allele variants remained undetected. However, this would only add more branches to the phylogenetic tree, while the current branching order would still be the same, making it of little relevance to the present study.

Four of the 75 Dutch samples were shown to contain *C. trachomatis* strains from both a non-LGV causing serovar as well as the LGV serovar L2. These samples were from heterosexual women, and in the western world LGV is almost always found among men who have sex with men. LGV serovars are invasive and easily cross contaminated. The four confirmed double infections are probably due to cross contamination with a L2 strain sometime during propagation in cell culture.

4.2 Unmatched discriminatory capacity of the MLST system and new sequence variation

Once again the MLST system proved its unmatched discriminatory capacity. Out of the 75 Dutch samples it managed to discriminate 47 different MLST genotypes, as compared to 18 genotypes for the traditional *ompA* genotyping. This shows the great potential of the MLST system for high resolution molecular epidemiology.

The Dutch samples contained in total 23 new allele variants. Most of these were found in the *hctB* and *pbpB* region, adding even more to the unrivalled variability of those regions. The *hctB* region now consists of nine distinct length variants, as compared to four when the system was presented by Klint *et al*. in 2007. Thirty-eight out of the 47 Dutch MLST profiles were completely new, or in other words, only nine were previously in the database. Since most of the previous data in the MLST database came from Swedish samples, this indicates that the *C. trachomatis* populations in Sweden and the Netherlands are quite distinct. There is probably lots of new sequence variation left to be discovered in samples from other countries.

The reason why no new allele variants were found when genotyping the French samples probably is that those samples are from men who have sex with men. It supports what has

already been suspected, i.e. that men who have sex with men are few in comparison to heterosexuals and seldom use condoms, have a high number of partners and also more sexual contacts abroad, making these types of sexual networks more global and less diverse with respect to the number of different *C. trachomatis* strains circulating (Klint *et al*,. 2006).

4.3 Optimization: Removal of redundant *hctB* allele variants and shortening of the *pbpB* target region

The manual alignment and removal of redundant *hctB* allele variants revealed that samples had been incorrectly genotyped in previous work at the department, which resulted in different MLST profiles for samples that in fact were identical. This has been corrected. One additional *pbpB* allele variant was found when re-amplifying 25 of the Dutch samples as longer 2366 base pair fragments instead of two shorter fragments. In total six out of 44 allele variants would be lost due to a shortening. However, the *pbpB* region was shortened anyways, considering the already high resolution of the MLST system and that the *pbpB* region is the most variable of the five target regions. This change will increase the number of samples that are successfully genotyped.

4.4 Future of the MLST system

The MLST database has grown significantly and an on-going study in which trachoma samples are being sequenced will hopefully add even more allele variants. Therefore the whole MLST system should be reviewed on a more statistical and bioinformatical level to simplify further optimization.

The MLST system for *C. trachomatis* is still quite new though. It is currently only in use in Uppsala and only a limited set of clinical samples has been genotyped so far. The MLST system needs a comprehensive database to take advantage of its full potential, and therefore it is of interest to spread the use of the system to other research groups around the world. Many MLST systems for other organisms utilize common online databases (such as http://pubmlst.org/). Perhaps the MLST database for *C. trachomatis* should be adapted to be put online to encourage others to use it and to make sure that all future information about allele variants and genotype profiles are efficiently collected in one single database. When the MLST system has been optimized and most of the genetic variation of the target regions has been gathered, then it might be fruitful to change the format.

At present the MLST target regions are being amplified individually. In a multiplex PCR several regions are amplified in one reaction, using more than one primer pair. If some, or all, of the MLST target regions could be multiplexed it would be a great improvement. Multiplexing a PCR requires a lot of optimization though, which can be very tedious and time-consuming. Besides, the bacterial DNA usually is of very low abundance in clinical samples, so high sensitivity is needed. Maybe new multiplex-ready PCR kits, which are already partly optimized, could help. And perhaps one could utilize real-time PCR to increase the sensitivity. An alternative option could be to amplify only the strain specific single nucleotide polymorphisms (SNPs) using a MegaPlex PCR, instead of the entire regions. In a MegaPlex PCR up to 75 SNPs can be amplified in one single reaction (Meuzelaar *et al.*, 2007).

Another interesting possibility is to put the system on a DNA microarray. All allele variants for all regions are then printed onto a chip and used as probes. The sample DNA is amplified in either single or multiplex PCR using fluorescently labeled primers and then transferred to the DNA chip where the products would hybridize to the immobilized probes. The light emitted from the fluorescent primers, incorporated into the hybridized products, is collected and analysed. If the hybridization is a perfect match the light intensity is higher then if it contains mismatches. The genotypes can be assigned based on the ratios of the fluorescence intensities. Having the MLST system on a DNA microarray would avoid the need to sequence the samples, which would save a lot of time.

Or one could use a liquid microbead array, e.g. Luminex xMAP® technology, in which sets of microbeads have unique spectral signatures created by coating the beads with various concentrations of two fluorescent dyes (Ugozzoli, 2004). In this approach, probes representing all of the MLST allele variants are attached to a unique set of microbeads each. The sample DNA is amplified in either single or multiplex PCR using primers labeled with a third fluorescent dye. When added to the reaction vessel the sample DNA would hybridize to its allele specific probe to a higher extent than to other probes. The microbeads and their probe-DNA duplexes are then detected and quantified using flow cytometry. Only products containing all three fluorescent dye are counted, avoiding problems with primer dimers and probe-probe interactions (Ugozzoli, 2004). As with the DNA microarray, using a liquid microbead array would avoid having to sequence the samples and save a lot of time.

At this time the MLST system for *C. trachomatis* is mostly suitable for research. If it is to be viable in routine molecular epidemiology and contact tracing it needs to be simplified and have its throughput increased, but without decreasing specificity or sensitivity. Perhaps multiplex PCR amplification combined with DNA array technology can achieve these goals.

5. Materials and Methods

5.1 Patient materials

Seventy-five *C. trachomatis* strains were isolated in the Netherlands between 2000 and 2005 from heterosexual women. The strains have been propagated in eukaryotic cell cultures. Information concerning the clinical symptoms expressed by the patients is available and were therefore suitable to use when exploring a possible correlation between multilocus sequence typing (MLST) genotypes and clinical symptoms. The samples were kindly provided by Servaas Morre, VU University Medical Center, Amsterdam.

Thirty samples were collected in France between 2005 and 2007 from men who have sex with men as part of a bigger multi-centre study. Fifteen of the samples are non-LGV serovars and the other 15 are LGV causing serovars. The samples were kindly provided by Bertille de Barbeyrac, Victor Segalen Bordeaux 2 University, Bordeaux.

Samples collected in Örebro in 2006 were kindly provided by Margaretha Jurstrand, Örebro Medical Centre Hospital, Örebro.

5.2 DNA extraction

DNA was purified from *C. trachomatis* cultures using a MagAttract DNA Mini M48 kit (QIAGEN, Hilden, Germany) on a BioRobot M48 workstation (QIAGEN), according to the manufacturer's instructions.

5.3 PCR amplification

The MLST target regions were amplified using the Expand High Fidelity PCR system (Roche Applied Science, Mannheim, Germany). This mix contains a polymerase with proofreading abilities which minimizes the risk of introducing novel mutations during the PCR amplification.

The PCR reactions, except for amplification of the *ompA* fragment, contained 7 µl of purified *C. trachomatis* DNA in a total volume of 25 µl, including 0.4 µM of each primer (Table 7), 0.2 µM dNTP, 1.5 mM $MgCl_2$ and 1.3 U Expand High Fidelity polymerase (Roche Applied Science, Mannheim, Germany). Cycling conditions were initial denaturation at 94 °C for 2 min, followed by 40 cycles of denaturation at 94 °C for 15 s, annealing at 60 °C for 30 s and elongation at 72 °C for 1 min. The elongation step was increased by 5 s per cycle after 10 cycles. The amplification was terminated with elongation at 72 °C for 7 min followed by indefinite cooling at 4 °C. The length of the amplified regions varied between 378 and 1499 base pairs. The longer 2366 base pair pbpB fragment used different cycling conditions. The annealing temperature was initially 68 °C, but it was decreased by 1 °C per cycle during the first 10 cycles down to 58 °C. Elongation was 80 s long the first 10 cycles, but it was then increased by 5 s per cycle for the remaining 30 cycles.

The *ompA* fragment was amplified by nested PCR using the HotStar Taq DNA polymerase (Qiagen, Hilden, Germany). In the first PCR 10 µl of purified *C. trachomatis* DNA was added

to a final volume of 40 µl. In the second PCR 2 µl of sample, from the initial run, was added to a final volume of 40 µl. The PCR reactions contained 0.4 µM of each primer (Table 7), 0.2 mM dNTP, 2.5 mM $MgCl_2$ and 1.5 U HotStar *Taq* DNA polymerase (Qiagen, Hilden, Germany). Cycling conditions were initial denaturation at 95 °C for 15 min, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s and elongation at 72 °C for 90 s. The amplification was terminated with elongation at 72 °C for 5 min followed by indefinite cooling at 4 °C.

All PCR amplifications also used water as a negative control and previously successfully amplified samples as positive controls.

Table 7. Primers used for PCR amplification and sequencing

Region	Name	Function	Sequence (5'→ 3')
CT058	CT222F	PCR, sequencing	CTTTTCTGAGGCTGAGTATGATTT
	CT1678R	PCR, sequencing	CCGATTCTTACTGGGAGGGT
	CT811F	Sequencing	CGATAAGACAGATGCCGTTTTT
	CT1022R	Sequencing	TAAGCACAGCAGGGAATGCA
CT144	CT144:248F	PCR, sequencing	ATGATTAACGTGATTTGGTTTCCTT
	CT144:1046R	PCR, sequencing	GCGCACCAAAACATAGGTACT
CT172	Four268F	PCR, sequencing	CCGTAGTAATGGGTGAGGGA
	Four610R	PCR, sequencing	CGTCATTGCTTGCTCGGCTT
hctB	hctB39F	PCR, sequencing	CTCGAAGACAATCCAGTAGCAT
	hctB794R	PCR, sequencing	CACCAGAAGCAGCTACACGT
pbpB-1	pbpB1F	PCR, sequencing	TATATGAAAAGAAAACGACGCACC
	pbpB823R	PCR, sequencing	CAGCATAGATCGCTTGCCTAT
pbpB-2	pbpB1455F	PCR, sequencing	GGTCTCGTTTTTGATGTTCTATTC
	pbpB2366R	PCR, sequencing	TGGTCAGAAAGATGCTGCACA
pbpB 2366	pbpB46F	Sequencing	GCTAACAGACTACTTGTGGGAT
base pair	pbpB785F	Sequencing	GTGTAGAAGCTTTTTGTGATCGTA
fragment	pbpB1377F	Sequencing	GCTTTCTATTTATCTGCAAGGAAC
	pbpB941R	Sequencing	GCTTGTAGCTCTGTAGAGATC
	pbpB1556R	Sequencing	TCTCTTTCAGCAAGGATAGCTTTA
	pbpB2333R	Sequencing	GCAGATACTAACTTAAAAATAGACC
ompA	P1	1st PCR	ATGAAAAAACTCTTGAAATCGG
	OMP2	1st PCR	ACTGTAACTGCGTATTTGTCTG
	MOMP87	2nd PCR, sequencing	TGAACCAAGCCTTATGATCGACGGA
	RVS1059	2nd PCR, sequencing	GCAATACCGCAAGATTTTCTAGATTTCATC
	ctr200F	Sequencing	TTAGGIGCTTCTTTCCAATAYGCTCAATC
	ctr254R	Sequencing	GCCAYTCATGGTARTCAATAGAGGCATC

Real-time PCR was performed by Markus Klint as described in the literature (Chen *et al.*, 2008) to confirm suspected double infections with LGV causing serovars. This method utilizes two probes specific for LGV and non-LGV serovars respectively. If both probes give signals then the sample is double infected.

5.4 Gel electrophoresis

All PCR products were evaluated using 1 % w / v agarose gels containing 10 µg / ml ethidium bromide and TBE buffer pH 8.3 (1.2 mM EDTA Titriplex, 44.5 mM boric acid and 44.6 mM TRIS). Electrophoresis was performed at 4 V / cm for 40 min.

5.5 Sequencing

ExoSAP-IT (GE Healthcare, Uppsala, Sweden), a premade solution of Exonuclease I and Shrimp Alkaline Phosphatase, was used according to the manufacturer's instructions prior to sequencing PCR in order to purify the samples from the Netherlands and Örebro. When puryfing the samples from France a cheaper solution was used instead, made by mixing Exonuclease I (Fermentas) and FastAP Thermosensitive Alkaline Phosphatase (Fermentas) according to the manufacturer's instructions.

The sequencing PCR was performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. The primers that were used are listed in Table 8. Cycling conditions were initial denaturation at 96 °C for 60 s, followed by 30 cycles of denaturation at 96 °C for 10 s, annealing at 55 °C for 5 s and elongation at 60 °C for 3 min, and then indefinite cooling at 4 °C.

Samples for sequencing were precipitated by addition of 50 µl of 99.5 % ethanol directly to the 20 µl of product from the sequencing PCR followed by centrifugation at 3000 x g for at least 30 min. All centrifugation was done at 4 °C. The supernatant was removed and the samples were centrifuged inverted at 190 x g for 1 min to remove residual liquid. Seventy µl 70 % ethanol was added and the samples were centrifuged at 1650 x g for at least 30 min. The supernatant and residual liquid was once again removed as described above. The remaining ethanol was allowed to evaporate during a 10 min incubation at room temperature, after which the DNA was dissolved in 20 µl Hi-Di Formamide (Applied Biosystems, Foster City, CA).

Sequencing utilized an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. Some samples were sent to the Uppsala Genome Center for sequencing.

The sequencing data were analysed using the software Sequencing Analysis 5.2 (Applied Biosystems), BioEdit 7.0.9 (Ibis Therapeutics, Carlsbad, CA) and ContigExpress which is part of VectorNTIAdvance 10.3.0 (Invitrogen).

5.6 Phylogeny

The phylogentic analyses were made using the software BioEdit 7.0.9 (Ibis Therapeutics, Carlsbad, CA) and PAUP* 4.0b10 (Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts). BioEdit, utilizing ClustalW, was used to construct correct alignments and to export this data into nexus files, which can be read by PAUP*. A phylogenetic tree was created by PAUP* using the clustering method neighbor-joining, without testing the model. The tree was bootstrapped 10 000 times, to get an estimate of how reliable the observed tree was with respect to the branching order of the different clades. The tree was rooted by using a corresponding sequence from the mice and hamster pathogen *C. muridarum* as an outgroup.

6. Acknowledgments

I would like to thank Björn Herrmann for the opportunity to do this exciting project. I greatly appreciate his enthusiasm and continuous support. And I would like to thank Markus Klint for guiding me through the techniques and for always answering my questions. I also want to thank my colleagues and everyone else at Mikrobiologen for being nice and friendly. Thank you!

7. References

Belland R., Ojcius D.M., Byrne G.I.. 2004. Chlamydia. *Nature Reviews Microbiology* **2**: 530-531.

Brunelle B.W., Sensabaugh G.F. 2006. The ompA gene in Chlamydia trachomatis differs in phylogeny and rate of evolution from other regions of the genome. *Infection and Immunity* **74**: 578-85.

Bush R.M., Everett K.D. 2001. Molecular evolution of the Chlamydiaceae. *International Journal of Systematic Bacteriology* **51**: 203-220.

Byrne G.I., Ojcius D.M. 2004. Chlamydia and apoptosis: life and death decisions of an intracellular pathogen. *Nature Reviews Microbiology* **2**: 802-808.

Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance, 2006. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, October 2007.

Chen C.Y., Chi K.H., Alexander S., Ison C., Ballard R.C. 2008. A real-time quadriplex PCR assay for the diagnosis of rectal lymphogranuloma venereum (LGV) and non-LGV Chlamydia trachomatis infections. *Sexually Transmitted Infections* [Epub ahead of print] doi: 10.1136/sti.2007.029058

Da Ros C.T., Schmitt Cda. S. 2008. Global epidemiology of sexually transmitted diseases. *Asian Journal of Andrology* **10**: 110-114

Dautry-Varsat A., Balañá M.E., Wyplosz B. 2004. Chlamydia – host cell interactions: recent advances on bacterial entry and intracellular development. *Traffic* **8**: 561-570

Elwell C.A., Ceesay A., Kim J.H., Kalman D., Engel J.N. 2008. RNA interference screen identifies Abl kinase and PDGFR signaling in Chlamydia trachomatis entry. *PLoS Pathogens* **4:** e1000021.

Everett K.D., 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. *International Journal of Systematic Bacteriology* **49**: 415-440.

Gomes J.P., Bruno W.J., Nunes A., Santos N., Florindo C., Borrego M.J., Dean D. 2007. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. *Genome Research* **17**: 50-60

Grieshaber N.A., Sager J.B., Dooley C.A., Hayes S.F., Hackstadt T. 2006. Regulation of the Chlamydia trachomatis histone H1-like protein Hc2 is IspE dependent and IhtA independent. *Journal of Bacteriology* **188**: 5289-5292.

Hafner L.M., McNeilly C. 2008. Vaccines for Chlamydia infections of the female genital tract. *Future Microbiology* **3**: 67-77.

Herrmann B., Pettersson B., Everett K.D., Mikkelsen N.E., Kirsebom L.A. 2000. Characterization of the *rnpB* gene and RNase P RNA in the order Chlamydiales. *International Journal of Systematic and Evolutionary Microbiology* **1**: 149-158.

Herrmann B. 2007. A new genetic variant of Chlamydia trachomatis. *Sexually Transmitted Infections* **83**: 253-254

Jurstrand M., Falk L., Fredlund H., Lindberg M., Olcén P., Andersson S., Persson K., Albert J., Bäckman A. 2001. Characterization of Chlamydia trachomatis omp1 genotypes among sexually transmitted disease patients in Sweden. *Journal of Clinical Microbiology* **39**: 3915-3919.

Klint M., Löfdahl M., Ek C., Airell A., Berglund T., Herrmann B. 2006. Lymphogranuloma venereum prevalence in Sweden among men who have sex with men and characterization of Chlamydia trachomatis ompA genotypes. *Journal of Clinical Microbiology* **44**: 4066-4071.

Klint M., Fuxelius H.H., Goldkuhl R.R., Skarin H., Rutemark C., Andersson S.G., Persson K., Herrmann B. 2007. High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis. *Journal of ClinicalMicrobiology* **45**: 1410-1414.

Low N., Egger M., Sterne J.A., Harbord R.M., Ibrahim F., Lindblom B., Herrmann B. 2006. Incidence of severe reproductive tract complications associated with diagnosed genital chlamydial infection: the Uppsala Women's Cohort Study. *Sexually Transmitted Infections* **82**: 212-218

Lysén M., Osterlund A., Rubin C.J., Persson T., Persson I., Herrmann B. 2004. Characterization of ompA genotypes by sequence analysis of DNA from all detected cases of Chlamydia trachomatis infections during 1 year of contact tracing in a Swedish County. *Journal of Clinical Microbiology* **42**: 1641-1647.

Mabey D.C., Solomon A.W., Foster A. 2003. Trachoma. *Lancet* **362**: 223-9. Review.

Maiden M.C. 2006. Multilocus sequence typing of bacteria. *Annual Review of Microbiology* **60**: 561-588.

Miller K.E. 2006. Diagnosis and treatment of Chlamydia trachomatis infection. *American Family Physician* **73**: 1411-1416.

Meuzelaar L.S., Lancaster O., Pasche J.P., Kopal G., Brookes A.J. 2007. MegaPlex PCR: a strategy for multiplex amplification. *Nature Methods* **4**: 835-837.

Morrison R.P. 2003. New insights into a persistent problem – chlamydial infections. *The Journal of Clinical Investigation* **111**: 1647-9. Review.

Moulder J.W. 1966. The relation of the psittacosis group (Chlamydiae) to bacteria and viruses. *Annual Review of Microbiology* **20**: 107-130

Mpiga P., Ravaoarinoro M. 2006. Chlamydia trachomatis persistence: An update. *Microbiological Research* **161**: 9-19.

Ripa T., Nilsson P.A. 2006. A variant of Chlamydia trachomatis with deletion in cryptic plasmid: implications for use of PCR diagnostic tests. *Euro Surveillance* **11**: E061109.2

Ripa T., Nilsson P.A. 2007. A Chlamydia trachomatis strain with a 377-bp deletion in the cryptic plasmid causing false-negative nucleic acid amplification tests. *Sexually Transmitted Diseases* **34**: 255-256.

Soderblom T., Blaxhult A., Fredlund H., Herrmann B. 2006. Impact of a genetic variant of Chlamydia trachomatis on national detection rates in Sweden. *Euro Surveillance* **11**: E061207.1

Swedish Institute for Infectious Disease Control. 2008. Statistik för klamydiainfektion. WWW document March 2008:
http://www.smittskyddsinstitutet.se/statistik/klamydiainfektion/ Date visited: 6 May 2008.

Taylor H.R. 2008. Trachoma: A Blinding Scourge from the Bronze Age to the Twenty First Century. Centre for Eye Research Australia, Melbourne.

Ugozzoli L.A. 2004. Multiplex assays with fluorescent microbead readout: a powerful tool for mutation detection. *Clinical Chemistry* **50**: 1963-1965.

Thygeson P. 1939. The Treatment of Trachoma with Sulfanilamide: A Report of 28 Cases. *The American Ophthalmological Society* **37**: 395-403.

World Health Organization. 2007. Trachoma. WWW document 22 June 2007:
http://www.who.int/blindness/causes/trachoma/en/index.html Date visited: 6 May 2008.

Wu I.B., Schwartz R.A. 2008. Reiter's syndrome: The classic triad and more. *Journal of the American Academy of Dermatology* [Epub ahead of print] doi:10.1016/j.jaad.2008.02.047

The front page picture shows a human pap smear with *C. trachomatis* in the vacuoles at 500 times magnification and stained with hematoxylin and eosin. It was taken in 1988 at Dr. Lance Liotta's laboratory, US, and has been released into the public domain.

8. Appendix

Table 8. Complete color coded MLST profiles of samples from patients with clinical symptoms.

Sample ID	CT058	CT144	CT172	*hctB*	*pbpB*	*ompA*	SE genotype
N1	8	1	7	8	18	30	B
N2	8	1	7	8	18	30	B
N4	12*	7	2	12	10	1	D
N5	26	7	2	5	40	31	D
N6	19	7	2	5	37	31	D
N7	27	18	2	7	34	1	D
N8	19	7	2	5	37	31	D
N9	19	7	1	5	37	1	D
N10	4	7	1	5	4	1	D
N11	4	1	4	10	23	32	D
N12	4	1	4	34	23	2	D
N13	4*	1	3	10	23	2	D
N14	4*	9	3	10	23	2	D
N15	8	1	4	10	23	2	D
N16	8	1	4	10	23	2	D
N17	8	1	4	10	21	2	D
N18	13	19	6	18	28	22	L2
N19	28	7	2	5	37	33	D
N20	8	7	2	5	4	34	D
N21	2	6	14	1	2	6	E
N22	2	6	2	35	2	6	E
N23	19	6	2	5	41	6	E
N24	19	7*	2	1	1	6	E
N26	2	6	2	35	2	6	E
N27	19	7	3	5	1	6	E
N28	2	6	14	1	2	6	E
N29	19	7	2	7	1	6	E
N30	19	7	1	5	4	24	F
N31	19	7	1	5	4	24	F
N32	19	7	1	5	4	24	F
N34	13	19	6	18	28	22	L2
N35	19	7	1	5	4	24	F
N36	19	1	1	5	4	24	F
N37	19	7	1	5	4	24	F
N38	19	5	2	5	4	24	F
N39	6	10	1	10	6	9	G
N40	6	10	1	10	6	9	G
N41	6	10	1	10	6	9	G
N42	19	1	1	10	6	11	G
N43	8	1	4	10	42	11	G
N44	5	12	3	10	5	11	G
N45	8	1	3	10	5	11	G
N46	6	10	1	10	6	9	G
N47	8	1	3	10	5	11	G
N48	8	1	3	10	5	11	G
N49	8	1	4	10	21	18	H
N50	29	11	9	12	21	35	H
N51	12	11	9	12	21	35	H
N52	15	7	1 (??)	36	43 (?)	?	

N53	12	11	9	12	21	35	H
N54	29	11	9	12	21	35	H
N55	15	7	1	37	43	38	J
N56	12	11	9	12	21	35	H
N57	15	7	1	36	43	38	J
N58	5	12	7	10	45	36	I
N59	5	12	7	38	45	36	I
N60	13	19 (?)	6	18 (?)	28 (?)	22	L2
N61	5	12	7	10	45	36	I
N63	5	12	7	38	45	36	I
N66	30	1	7	10	5	37	I
N67	6	1	7	8	5	37	I
N69	5	12	7	10	45	36	I
N70	5	12	7	10	45	36	I
N71	5	12	7	10	45	36	I
N72	6	1	7	8	5	37	I
N73	13	19 (?)	6	18 (?)	28 (?)	22	L2
N76	12	11	9	12	22	20	J
N78	4	1	1	10	22	20	J
N81	8	9	4	10	8	12	K
N82	4	1	7	10	44	12	K
N83	7	20	4	33	8	12	K
N84	7	9	4	10	8	12	K
N85	7	9	3	10	8	12	K
N86	7	9	1	10	8	12	K
N87	12	21	9	12	24	12	K

The numbers correspond to allele variants in the MLST database. The different MLST profiles are color coded to make it easier to see which ones are identical. White profiles are unique. An asterisk (*) indicates sequences of somewhat lower quality. This has been assessed to have no impact on the results and conclusions (Discussion section 4.1). SE genotype is the genotype attained from *ompA*-genotyping. A question mark means that there are double peaks on strain specific nucleotides, suggesting a possible double infection. Samples N18, N34, N60 and N73 have been confirmed with real time PCR to be double infected with a non-LGV causing serovar and the LGV serovar L2. Sample N52 is believed to be double infected with two non-LGV causing serotypes.

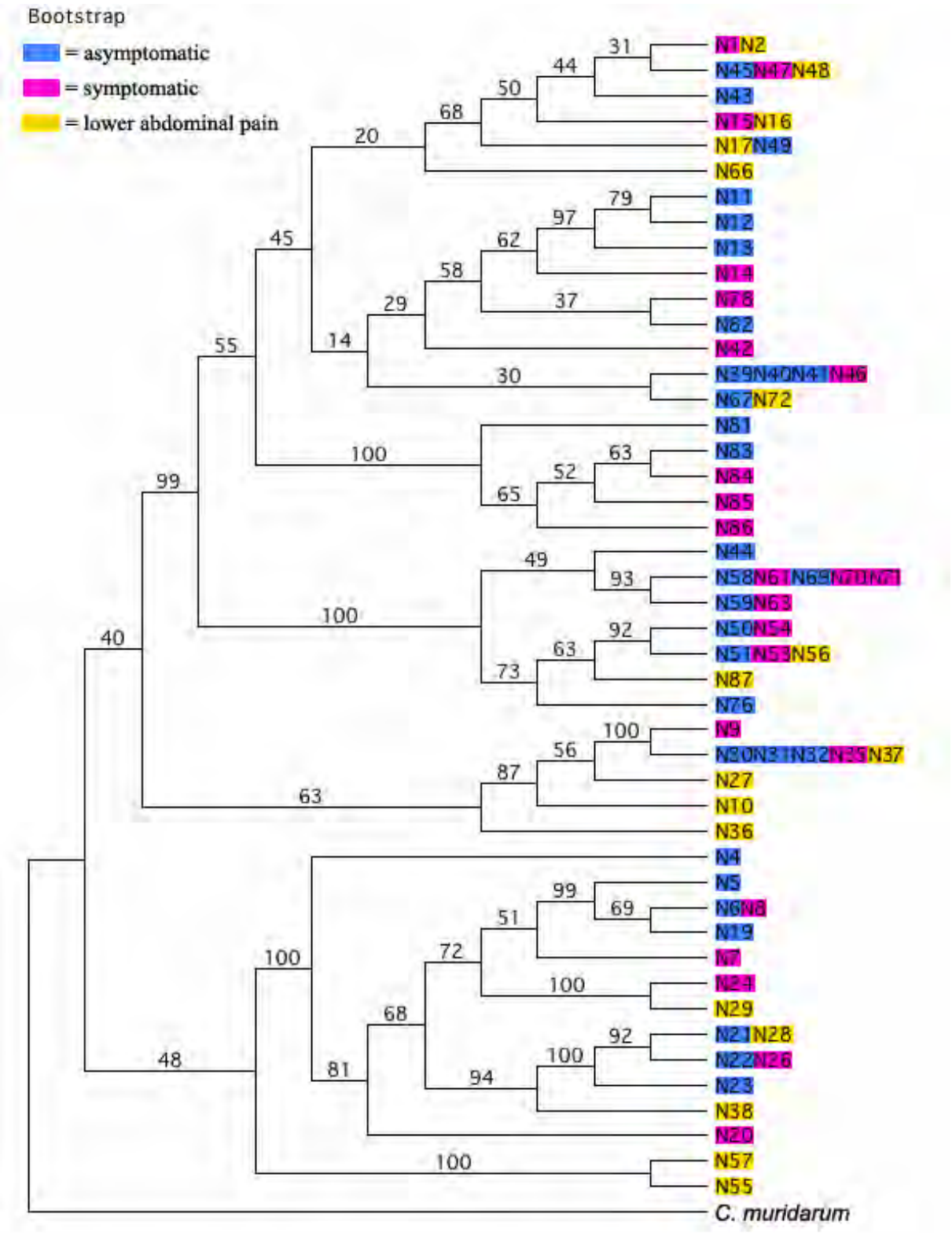


Figure 6. A cladogram showing the schematic genetic relationship between the Dutch samples as well as the reliability of the branching order. The tree was bootstrapped 10 000 times and the reliability is displayed as the percentage of those 10 000 replicates in which a certain clade appeared. Roughly one can be quite certain in a clade appearing more than 90 % of the time, and not very certain in clades appearing less than 50 % of the time. Blue, pink and yellow represent samples from asymptomatic and symptomatic patients and from patients with lower abdominal pain, respectively. The tree has been rooted with a corresponding sequence from *C. muridarum* as an outgroup.

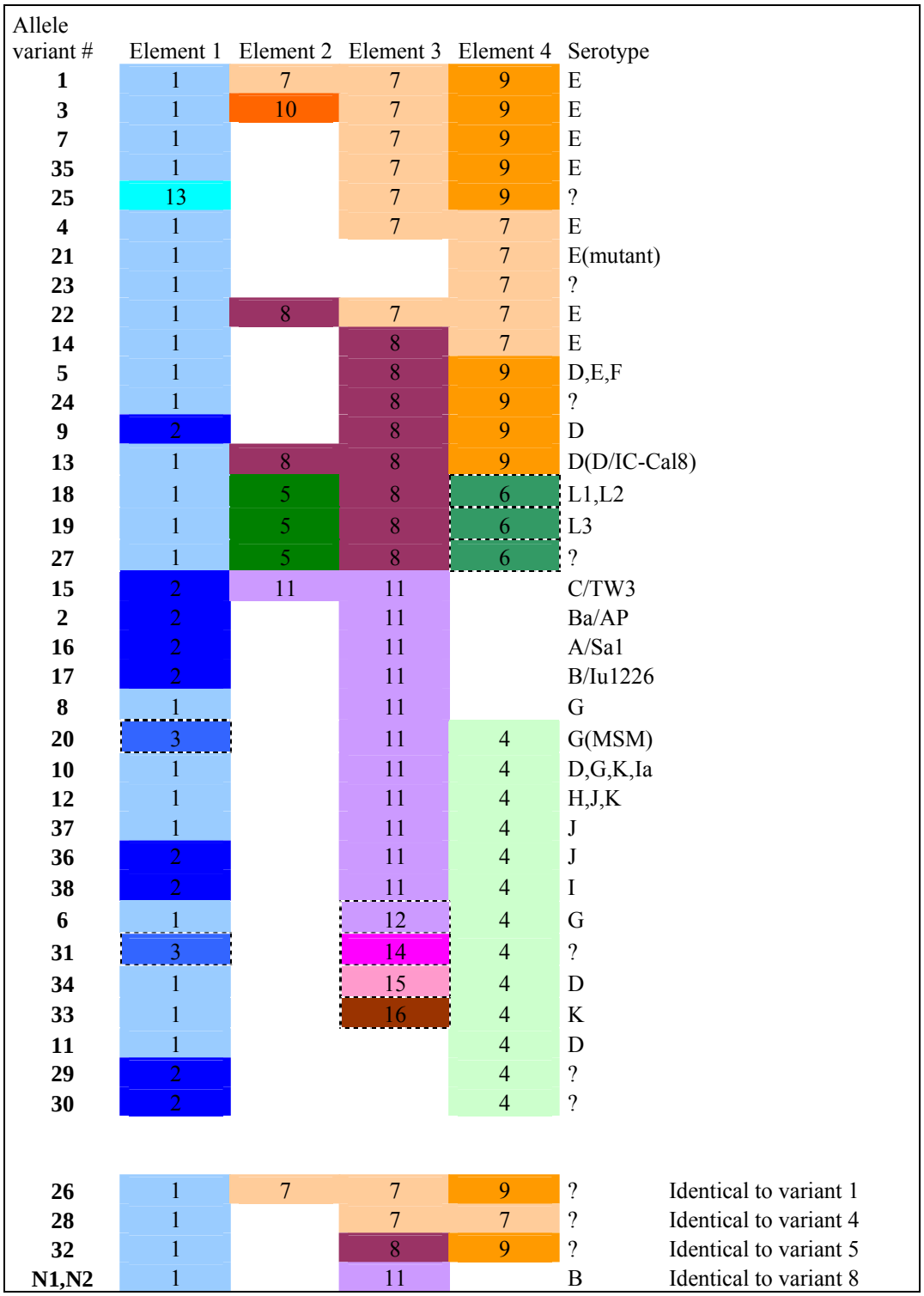


Figure 7. Schematic view of the larger repetitive elements of all the allele variants in the *hctB* sequence database. The different colors and their numbers refer to different types of repetitive elements. Elements marked with a striped line are not full length, i.e. not 108 nucleotides long. Variants which have identical repetitive elements can still have single point mutations outside of the repetitive region, mutations which make them unique. However, that is not the case for the four variants which have been marked as identical and redundant. N1,N2 was first given the allelic number 33, but this number was reassigned since all samples which were incorrectly genotypes as 33 was changed to 8.