Molecular epidemiology of Chlamydia trachomatis

Bom, R.J.M.

Citation for published version (APA):
Bom, R. J. M. (2014). Molecular epidemiology of Chlamydia trachomatis

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
6.
General discussion
Chlamydia trachomatis infections are a large public health and economic burden. Typing methods that discriminate between clinically, biologically or epidemiologically different C. trachomatis strains, can improve screening and prevention interventions to decrease the burden caused by this pathogen. The studies in this thesis deal with the development and evaluation of such typing methods. In addition, transmission patterns of the different C. trachomatis strains discerned by the typing methods were investigated in large epidemiological studies. The implications of our findings are discussed in the context of recent literature.

As lymphogranuloma venereum (LGV) inducing C. trachomatis strains require a prolonged regimen of antibiotics compared to other C. trachomatis strains, distinction between LGV and non LGV-inducing strains is critical for good clinical management and to prevent ongoing transmission. In Chapter 2, we evaluated the diagnostic performance of a newly developed real time PCR that discriminates between these strains based on a deletion in the pmpH gene. This new typing assay had the same sensitivity for LGV-inducing strains as reverse hybridisation or ompA sequencing techniques, for no LGV-inducing strains were missed in the studies using this real time PCR. The sensitivity of the assay did not decrease in the presence of a concomitant non LGV-inducing strain. In addition of being faster and less laborious than other typing methods, real-time PCR provides a score, the cycling threshold, which indicates the quality and reliability of the test result by indicating the amount of chlamydial genomic DNA within the sample.

As all amplification assays, including sequencing analysis, rely on the initial amount of DNA within the samples, the quantity of this initial amount of DNA is indicative for the success of the typing procedure. This real-time PCR was used throughout our studies as the genomic DNA quantification method. With it, we could anticipate the outcome of our typing-by-sequencing assays, either through the adaptation of the cycling conditions to the amount of genomic DNA, or through the exclusion of samples when a successful outcome was unlikely. This reduced the work load and costs of our studies considerably, as we did not have to check the success of amplification for every target region for sequencing for each sample by means of gel electrophoresis.

Another possible advantage of this real-time PCR is that the bacterial load of a C. trachomatis infection can be estimated by adding a target for human genomic DNA to the assay. Since we used urine or swab samples, it should be noted that this load is not as precise as loads discerned from body fluids, such as plasma or serum. Therefore, estimated loads are likely to be influenced by differences between anatomical sites and may vary due to differences in urine
volume, secretions, menstrual cycle, or the number of bacteria swabbed. However, when obtained under similar sampling conditions or in cell cultures, results may be comparable.¹

In Chapter 3, we investigated two high resolution typing methods for *C. trachomatis* with a degree of resolution needed for molecular epidemiological studies: the multilocus sequence typing (MLST) method published by Klint et al. and the multilocus variable number of tandem repeat (VNTR) analysis (MLVA) published by Pedersen et al.²³ The MLST method included five variable regions (*hctB*, CT058, CT144, CT172, and *pbpB*) to which *ompA* was added, and the MLVA method combined *ompA* typing with analysing three highly variable single nucleotide repeats (CT1291, CT1299, and CT1335). Both methods were transformed into nested assays to increase the sensitivity needed for application on direct patient samples. Also, the target lengths of the sequence products were adjusted so that they could be analysed by sequencing in a single run. Both typing methods greatly increased the resolution of *C. trachomatis* typing compared to *ompA* sequencing alone and met guidelines set for molecular epidemiological studies.⁴ Although the methods were comparable in resolution, the data of the MLVA method became ambiguous when repeat lengths increased, which made it difficult to interpret the results in a consistent manner. The adjusted MLST procedure was therefore selected as our method of choice for molecular epidemiological studies.

The data acquired by the MLST method were depicted in minimum spanning trees to visualise the chlamydial population structure. These minimum spanning trees showed that *C. trachomatis* strains were distributed over multiple clusters that varied in size and heterogeneity. Although these clusters were often congruent with *ompA* genovars, in many cases identical genovars were separated into distinct clusters, and some clusters were composed of multiple *ompA* genovars. This reshuffling of genovars is caused by horizontal genetic exchange of the *ompA* gene. The MLST method is able to unveil important information on the population structure of *C. trachomatis*, which was previously missed by *ompA* genotyping. By determining these clusters from the minimum spanning trees, a distinction could be made between the various *C. trachomatis* strains in a more realistic fashion, which is very useful for molecular epidemiological analyses.

Although the MLST performs well in terms of resolution and sensitivity, the choice of targets has been criticised.⁵ While most MLST systems make use of canonical single-nucleotide polymorphisms (SNPs) in conserved housekeeping genes, our MLST system makes use of the most variable regions within the *C. trachomatis* genome. The use of housekeeping genes is preferred when sufficient diversity in a certain pathogenic genome is present. In case of highly conserved organisms like *C. trachomatis*, the use of genetic regions exhibiting higher diversity, such as genes under positive selection, is preferred over the inclusion of a larger number of low
polymorphic MLST loci. One should bear in mind that MLST was designed as a pragmatic typing technique that exploits population genetic analysis. In addition, the genetic population structure of an asexual organism is a result of clonal genetic transfer (mutation) that is spread by inheritance, as well as nonclonal genetic transfer (recombination), which is established through the horizontal exchange of genetic information. By selecting only diversity that has arisen from mutation, such as synonymous substitutions in housekeeping genes, the genetic population structure of the organism would be skewed towards a fully clonal paradigm. Recent studies, using the analysis of full genomes, have shown that horizontal gene transfer is also relatively common in C. trachomatis. Although mutation events happen more often during the natural history of C. trachomatis, the impact of recombination on the genome is larger due to the size of the recombined fragment. However, as more full genomes of C. trachomatis become available, the choice of targets for a typing system may be reconsidered to represent clonal and nonclonal genetic variation in a more optimal manner. In the future, it may be feasible to sequence the full genome, as rapid advances in sequencing techniques may enable us to acquire the full genome sequence of clinical samples with the same sensitivity, at the same costs, and in the same time as MLST nowadays. This would allow the analysis to be done later on, with any desired resolution, based on the preferred amount of clonal and nonclonal genetic variation, and in addition of certain genes of interest, distribution proved to be much more apparent (Table 1). Among MSM, the majority of urogenital C. trachomatis strains belonged to two large clusters: one consisted fully of genovar D samples, and the other consisted of both genovars G and J. Also, an additional smaller cluster of genovar D type infections was found among MSM. The few residual samples were mainly genovars E and F, and were found in smaller clusters or as singletons. Finally, a cluster of LGV-inducing types existed that were genetically more diverse than previously reported in Europe. A recent publication, however, reported on the circulation of a second LGV-inducing strain among MSM that resembles the L2 ompA type strain found in our study.

The samples found among heterosexuals were located in multiple clusters of various sizes or as singletons within the minimum spanning trees. Two of the three largest clusters consisted fully of genovar E samples, while the third large cluster was composed of mainly genovar F samples, in addition to some samples with genovars D and J. In the remaining smaller clusters and singletons, all genovars were found.

When plotting the samples from heterosexuals and MSM in the same minimum spanning tree, there was very little overlap between the C. trachomatis samples from the sex groups. Almost no samples from heterosexuals were found in the MSM-associated clusters and the majority of genovar D, G, and J samples found in heterosexuals belonged to the multiple small clusters outside the MSM-associated clusters. However, a small
proportion of samples from MSM, mainly genovars E and F, were dispersed evenly over the clusters associated with heterosexuals (Table 1).

*C. trachomatis* strains circulating among MSM proved to be much less diverse than among heterosexuals, where a more heterogeneous strain composition was found. Minimum spanning trees showed that most samples from MSM were found in just a few clusters. Within these clusters, most samples belonged to one or a few sequence types, with the remaining types closely related to them. Among heterosexuals many more clusters of various sizes and singletons were found. Within the clusters much more heterogeneity was observed: many more sequence types were present that could vary in up to 3 loci from the central sequence types (Table 1).

In Chapter 4 & 5, we used our MLST system to explore epidemiological differences for the various *C. trachomatis* strains within MSM and heterosexual populations. As was shown previously, MSM infected with LGV-inducing strains form a distinct subpopulation. These men were characterised by much higher sexual risk behaviour and higher rates of other STIs, including HIV, compared to other MSM. However, no associations were found between the urogenital strains and age, ethnicity, lifestyle, partner characteristics, or sexual behaviour among the MSM population in Amsterdam, the Netherlands. In addition, various men were infected with different *C. trachomatis* strains at different visits; with different strains at different anatomical locations at the same visit; or even with different strains at the same anatomical location at the same visit. The distribution of these urogenital strains found in the Netherlands

<table>
<thead>
<tr>
<th></th>
<th>MSM population</th>
<th>Heterosexual population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most predominant <em>ompA</em> genovars</td>
<td>D, G, J</td>
<td>E, F, D</td>
</tr>
<tr>
<td>Number and size of risk group-specific clusters</td>
<td>A few large clusters</td>
<td>Multiple clusters of various size</td>
</tr>
<tr>
<td>Diversity within the risk group-specific clusters</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Overlap with other risk group-specific clusters</td>
<td>A small proportion</td>
<td>Absent</td>
</tr>
<tr>
<td>Subpopulations within risk group-specific clusters</td>
<td>No</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>Geographic variation within risk group-specific clusters</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Table 1.* Differences between MSM and heterosexuals in *C. trachomatis* genotypes and population structures.
was very similar to the distribution found in Sweden and the United States, and the composition of strains did not seem to change over time. We therefore concluded that there was a co-occurrence of different *C. trachomatis* strains, endemic within the MSM population at large. However, when distinguishing MSM-associated strains from heterosexual-associated strains, we did find some associations. MSM who were younger and MSM who reported sex with women were more often infected with heterosexual-associated strains. These strains were also more often found among MSM in the United States, compared to the Netherlands and Sweden (Table 1).

Among heterosexuels, differences in the composition of *C. trachomatis* strains were observed between different subpopulations within one country. Although native Dutch and Surinamese migrants in the Netherlands were both represented in all clusters, the proportion of each ethnic group differed significantly between the clusters. In Paramaribo, Suriname, the various ethnic groups also differed in the distribution of *C. trachomatis* strains. The same effect was seen when a comparison was made between various countries. The distribution of samples found in the Netherlands differed from the distributions found in Suriname (a country historically and culturally related to the Netherlands) and in Sweden (a country geographically close to the Netherlands). Here too, only the proportion of each country within each cluster differed. However, when the sample distribution found in the Netherlands was compared with the one found in China (a country both culturally and geographically distant to the Netherlands), clusters unique for one or the other country appeared, although also strains with identical MLST-patterns circulated in both countries. This clearly shows that within the heterosexual population variation in the distribution of *C. trachomatis* strains existed between culturally and geographically distinct groups, and that this variation increased with cultural and geographical distance (Table 1).

Within the Netherlands and Suriname, as well as between both countries, sexual mixing between different groups occurred frequently. Within the Netherlands, high levels of mixing were seen between the native Dutch and the Surinamese migrant populations. These Surinamese migrants also mixed with the native Surinamese population, forming a possible STI transmission bridge between the native populations of the Netherlands and Suriname. However, no clear evidence was found that strains associated with the native Dutch population were transmitted via the Surinamese migrants to the native Surinamese population, or vice versa. The distribution of *C. trachomatis* strains found among the Surinamese migrant population differed from both the native Surinamese distribution as well as from the Dutch distribution, but did not form an intermediate reflection of the native groups. In addition, when we examined the distribution of *C. trachomatis* strains among the participants that reported sexual mixing and those who did not, no significant differences could be found between the two
In the distribution of *C. trachomatis* clusters. In Paramaribo, almost half of the study population reported sexual mixing with other ethnic groups. Nonetheless, significant distribution differences of *C. trachomatis* strains were found between various groups. We hypothesised that the distinction in the distribution of *C. trachomatis* strains between ethnic groups may be explained by the absence of effective sexual mixing between the groups. As a consequence, transmission of *C. trachomatis* to other populations is limited and the prevalence of the clusters found among ethnic groups will vary due to stochastic effects.

In the studies described, we found that the differences between MSM and heterosexuals in *C. trachomatis* genotypes and population structures were much more extensive than different distributions of *ompA* genovars alone. Now, we will address the possible underlying causes behind these differences and make suggestions for future research on how to discern them. Most notably, MSM and heterosexual populations harbour distinct *C. trachomatis* strains. This can be explained by the near absence of sexual mixing between the two risk groups, as only a few men reported sex with both men and women, resulting in two separate transmission networks in which the specific strains reside. However, when we compare the distribution of strains found among heterosexuals in the Netherlands with the one found among heterosexuals in China, the differences are much smaller, even though we can assume that no transmission occurs between these two geographically and culturally distinct sexual networks.

As an alternative, we opt for the possibility of differences in tissue tropism between the risk group-specific strains. Differences in tissue tropism clearly exist between the various clades of *C. trachomatis* and within these clades, differences in tissue tropism may have evolved as well. As the pathogens reside in different niches within the human body (male urethral, cervical, and rectal tissue) and make use of different transmission routes (between male urethra and cervix, or between male urethra and anus), specialisation towards these niches or transmission routes may have occurred.

In our studies, we found no differences for urogenital strains in occurrence between urethral and rectal samples in MSM, or between urethral samples in heterosexual males and cervical samples in heterosexual females. As the transmission of *C. trachomatis* depends on the infection of the male urethra in both urogenital and anogenital transmission, differences in tissue tropism in risk group-specific *C. trachomatis* strains are likely to be found in the affinity to infect cervical or rectal tissue. These differences can be demonstrated by typing the *C. trachomatis*-positive samples from women, who have been tested both urogenitally and rectally, and who were found positive at one or both sites. From these data, the preference of a strain for cervical or rectal tissue may be deduced from the disproportional distribution of the strain over the two anatomical locations. Recent literature showed that rectal *C. trachomatis* infections occur in women in absence of anal intercourse by autoinoculation via cervical secretions.15,16
If some *C. trachomatis* strains have an increased preference for rectal tissue, this will influence the likelihood of auto-inoculation from cervix to rectum to occur. This is important for the testing procedures for *C. trachomatis* infections, as women are not routinely screened rectally in absence of a history of receptive anal intercourse and therefore these infections will remain undetected. The relation of a strain with a certain tissue can be reinforced by assessing the bacterial load of the strain and comparing this load with loads from other strains for each anatomical location. The bacterial load from a strain with a preference for a certain tissue will be higher on average, due to an increased proliferation in this specific tissue, although this will largely depend on the host’s immune response. Using bacterial loads also allows comparison between rectal samples from women and MSM, and urethral samples between heterosexual men and MSM.

These differences in tissue tropism may also be demonstrated in vitro, by assessing the ability of the MSM- and heterosexual-associated *C. trachomatis* strains to replicate in male urethral, cervical, and rectal tissue. The properties of these various tissues may be simulated by using polarised epithelial cell lines that are derived from these anatomical sites. So far, most cell lines used for culture do not resemble the tissues in which *C. trachomatis* naturally resides: the cells used are derived from non-epithelial tissues, or the epithelial cells used are not polarised. However, an endocervical cell line, the A2EN cell line, has recently been generated that closely mimics the in vivo cellular response to *C. trachomatis* infections. With the generation of proper urethral and rectal cell lines, an assay may be developed that can assess the ability to replicate in the three tissues for wild-type, naturally circulating *C. trachomatis* strains. This system may also be used to elucidate the differences in clinical symptoms of the current L2b-outbreak among MSM compared with the symptoms seen in classical LGV cases. In the current outbreak LGV seems to be only expressed as proctitis, while genital ulcerations and the typical inguinal buboes are very rare.

Another typical feature of MSM- and heterosexual-associated *C. trachomatis* strains is the difference in risk group-associated cluster variation. The heterosexual-associated strains are numerous and the heterogeneity found indicates a slowly evolving endemic disease that has diversified over time by stochastic effects. The MSM-associated strains, however, appear to be much more clonal and homogeneously clustered, with the divergent types closely related to the few predominant strains. They may therefore have arisen from more recent clonal outbreaks similar to the L2b outbreak, as they show resemblance to its shape, but in a more advanced state. If this is true, phylogenetic analysis on full genomes of MSM-associated strains will reveal a more common origin of these strains, compared to heterosexual-associated strains. Of special interest would be strains from MSM found outside the Western countries, like China. Analysis of these genomes may foster specific genomic adaptations towards MSM.
transmission networks. Such an analysis has been performed recently on a few clinical anorectal and cervical isolates, and three genes were discovered that were highly associated with rectal tropism in samples with \textit{ompA} genovar G.\textsuperscript{23} In our studies, we found that one of these genes, CT144, had a single variant that was present in the vast majority of samples derived from MSM, while being mostly absent in samples from heterosexuals. Interestingly, this allele was present in samples from nearly all MSM located in the two large MSM-associated clusters, as well as in samples from some MSM located in the heterosexual clusters.

The typical MSM- and heterosexual-associated \textit{C. trachomatis} clusters may also originate from network associated factors. Compared to heterosexuals, MSM in general, report higher numbers of partners. In addition, MSM mix more often with partners that differ in age, ethnicity, nationality, and lifestyle.\textsuperscript{24-26} Therefore the sexual network structure of MSM is much more interconnected, giving rise to a large international transmission network. Evidence for this can be seen in the rapid clonal spread of L2b among MSM throughout the world.\textsuperscript{27} These highly interconnected networks may be more susceptible to reduction of types through genetic drift, because of an increased genetic flow through the network, which may lead to fixation of only a few types. In contrast, less interconnected networks typically found among heterosexuals may lead to a more heterogeneous \textit{C. trachomatis} strain distribution. The reduced number of connections and non-random mixing of heterosexuals foster transmission networks that are smaller and have a more local spread. This can be seen from the so-called new variant \textit{C. trachomatis} outbreak: the strain was highly prevalent among heterosexuals in Sweden, but it has failed to spread successfully to other countries.\textsuperscript{28} Within these transmission networks the genetic flow is reduced and isolated sub-networks may exist. Within these sub-networks, strains may diversify and be more easily fixated due to the reduced effective population size. Full genome analyses may elucidate these genetic population structures as well as the phylogenetic tree would show deeper branches than in case of clonal expansion. If the MSM population can only harbour a few types, then the more heterogeneous heterosexual-associated strains found among MSM must have originated from the heterosexual population, although the majority of these strains were transmitted between men. It is unlikely to be solely due to bisexuals, as MSM-associated strains were almost completely absent within the heterosexual population, even though men infected with these strains reported more often sex with women. However, younger age was also related to being infected with a heterosexual-associated strain. As approximately one quarter of the MSM have their sexual debut with a woman, they may have contracted the infection at this period and later on introduced it in the MSM sexual network.\textsuperscript{25}

The effects of the structures of the different sexual networks on the \textit{C. trachomatis} strain distributions may be
resolved through mathematical modelling. In models containing parameters that influence the shapes of the transmission networks (such as number of partnerships, concurrency and non-random mixing), the influence of each of these parameters on the networks can be assessed independently. This way, the different population structures for various risk populations can be simulated, resembling the ones found in observational research. Next, these mathematical models can be used to determine the parameters that could potentially lower the prevalence of *C. trachomatis* within specific risk groups when targeted with tailored intervention measures. Molecular typing would then allow monitoring of these predicted changes experimentally within the population. With these monitoring efforts in mind, baseline studies performed in different countries are now being performed and data aggregation with subsequent public dispersion is offered via the *Chlamydia trachomatis* MLST database (mlstdb.bmc.uu.se). This database includes, at the time of writing, 2087 samples from various risk groups from 16 different countries from 6 continents. Enlarging this database will increase our knowledge of the distribution patterns of *C. trachomatis* that can be used for monitoring the spread of the infection in various sexual networks.

To assess the structures of the transmission networks for *C. trachomatis*, detailed epidemiological studies are needed that describe the determinants involved in the spread of the infection through a population, such as mixing patterns. These mixing patterns can be investigated using molecular typing techniques to assess whether successful transmission occurs. As seen in our study of Surinamese migrants, transmission of strains can be absent in the presence of sexual mixing, creating a subpopulation that is isolated from the general population. In conclusion, sexual mixing as reported by infected individuals seems less accurate to reveal possible transmission networks than molecular genotyping analysis combined with epidemiological data as presented in this thesis. In this way, epidemiological studies may determine the possible transmission networks of *C. trachomatis* and reveal the determinants that are important for the spread of the infection within and between these networks. Molecular typing studies can assess which strains are present, which dynamics they are subjected to, and the degree of mixing between various networks. Mathematical modelling may find the most effective transmission interventions for specific populations that could reduce the burden of infection. Integrating molecular typing of *C. trachomatis* with epidemiological research and mathematical modelling into one interdisciplinary research effort, could lead to better targeted interventions for people at risk for infection, and to prevention strategies resulting in maximum effect.

**References**


