

Chapter 7

Molecular Typing and Clustering Analysis as a Tool for Epidemiology of Infectious Diseases

Sylvia M. Bruisten and Leo Schouls

7.1 Introduction

This chapter describes the mechanism of typing procedures of human pathogens and gives some examples to substantiate the added value of typing and clustering analysis in epidemiology. Three steps need to be discerned in the process toward molecular clustering analysis. First, the pathogen must be recognized and identified (the diagnostic step). Second, the typing of the pathogen genome is performed and third, the clustering of the specific type with other known or newly identified types is needed for added epidemiological information.

For many infections, public health services monitor the transmission pattern of a disease and the possibility of outbreaks by performing active case finding (see Chapter 8). The ensuing data of both the index patients and their (possible) contacts are entered in local or central databases or in national registries. Traditionally, epidemiological data are obtained by subjecting patients to structured questionnaires – filled out by either themselves or by trained nurses or other public health workers. Studies of pathogen dissemination that rely solely on these self-reported behavioral data can yield incomplete or incorrect information. For example, for sexually transmitted infections (STI), the sensitive nature of the questions inhibits truthful answers. Nevertheless, identification of high-risk populations to whom prevention and intervention should be tailored is essential in infectious disease control. In the last decades, additional possibilities have emerged for contact tracing and for establishing patterns of transmissions of certain pathogens by using molecular typing techniques. Classical epidemiological data and the typing data should be analyzed together to interpret the data (Tenover et al. 1995). These typing tools work only for those cases in which the pathogen itself can be recovered from the infected person, which introduces a bias with respect to behavioral data. Molecular typing also requires genetic polymorphism within a species. Polymorphism will arise because pathogens adapt to different environments. For example, bacteria have

S.M. Bruisten (✉)
Cluster Infectious Diseases, Public Health Service of Amsterdam, Amsterdam, The Netherlands
e-mail: sbruisten@ggd.amsterdam.nl

to cross barriers for initial colonization, and subsequent survival requires specific virulence traits that are subject to variation. Characterization of the polymorphic traits and subsequent cluster analysis should reveal if distinct types or clusters of strains are associated with particular disease manifestations, for example, in the case of the “hamburger bacterium,” *Escherichia coli* type O157:H7 (Noller et al. 2006). Another example is the different serovars of *Chlamydia trachomatis* with diverse disease manifestations such as the ocular serovars A and B, the urogenital serovars C–F, and the lymphogranuloma-causing serovar L (Klint et al. 2007; Spaargaren et al. 2005; Hamill et al. 2007). Strains and clusters can, however, also be associated with distinct host risk groups, as was shown to occur for hepatitis A virus (Tjon et al. 2007). Typing may also aid to identify the source of an infection (person to person or environmental) and in case of a hospital (nosocomial) infection, typing can be used to assess preventative measures and treatment. In addition, typing may help to assess the effects of human intervention, such as antimicrobial treatment or vaccination, on the composition of the pathogen population.

Molecular methods are increasingly applied to aid the control of infectious diseases, both for detection and for typing of the pathogen. For detection purposes, the polymerase chain reaction (PCR) has a crucial role, but other nucleic acid amplification techniques are also increasingly used. Molecular typing and subsequent storage of this molecular data, as well as clinical and epidemiological data in libraries, can be helpful in identifying transmission routes within and among populations.

Typing is helpful in individual cases for

- a. aiding in source tracing to elucidate whether there is transmission from patient to patient or from an environmental source to patient
- b. revealing phenotypical properties of the pathogen: drug resistance, infectivity, virulence
- c. assessing treatment effectivity: relapse or possible reinfection
- d. identifying sampling and laboratory errors

Typing is helpful for public health issues:

- e. Libraries help to decide which isolates belong to a nosocomial or a community-acquired outbreak
- f. Vaccination issues: does the (live) vaccine strain cause disease or is it the wild-type pathogen?
- g. Intervention efficacy: do strains mutate and escape treatment or vaccination?

Epidemic and endemic clones need to be recognized and their geographical spread should be informative about the nature of the outbreak. Nowadays, several techniques for typing of pathogens are available. In this chapter, we introduce these typing methods and discuss their value for providing additional epidemiological information.

7.2 Attributes for Successful Typing

Typing methods not only should be practical but also need to be highly standardized to facilitate (inter)national comparison and data exchange. Ideally, typing methods should be easy to perform, rapid, reliable, highly discriminatory, and suitable for large-scale and widespread use (van der Zee et al. 1999). Reproducibility is also of great importance, because variation in typing results of samples repeatedly taken from the same person should ideally reflect biological variation due to different sampling sites or different time points of sampling, and not technical variation. Another important characteristic of typing is the discriminatory power that depends on both the variation of the typing marker used within the pathogen population and the typing procedure itself. Different formulas based on Simpson's index of diversity are in use to calculate this discriminatory power (Blanc 2004; van Belkum 2007).

Phenotypic typing methods such as biotyping (e.g., the ability to ferment certain sugars), serotyping (reactivity with type-specific antibodies), phage typing (sensitivity for bacteriophages), antibiogram typing (determining the resistance of a panel of antibiotics), and multilocus enzyme electrophoresis (MLEE) generally show less discriminatory power than do genotypic methods. In addition, the phenotypic characterization may vary due to variations in growth conditions. Genotypic methods encompass both fragment-based (i.e., PFGE, PCR-RFLP, AFLP, VNTR) and sequence-based methods (i.e., MLST, Spa typing, microarray).

In bacteria, and also in other microbes such as viruses, fungi or parasites, genetic variations can be divided into three categories (Arnold 2007):

- a. Small local changes in the nucleotide sequences (single nucleotide polymorphisms or SNPs)
- b. Intragenomic rearrangements (including insertions, deletions, and duplications)
- c. Acquisition of DNA sequences from other microorganisms or even from other living beings such as the host (horizontal gene transfer and recombination)

A small glossary is added at the end of this chapter. It contains definitions of terms used here.

7.3 Genotyping Methods

7.3.1 Plasmid Typing

One of the oldest methods to genotypically characterize bacteria is plasmid profiling (Snipes et al. 1989). Such methods have proven to be particularly useful if there is suspicion of dissemination of a particular resistance gene or set of genes (Singh et al. 2006). However, as plasmids are mobile genetic elements, they cannot be easily used in epidemiological linking studies.



Fig. 7.1 Bacterial culture to obtain an isolate (from Wikipedia website, http://en.wikipedia.org/wiki/Microbiological_culture). A culture of *Bacillus anthracis*

The fragment-based typing (FBT) techniques mostly apply to bacteria and can be subdivided into those that make use of sequence variation in the total genome and those that target particular genes or repeat elements. For both approaches it is usually required that the bacteria are first cultured on agar or in broth to start with pure cultures (Fig. 7.1). This will ensure that ample DNA is available. If the typing needs to be performed directly on patient samples, specific PCR amplification usually precedes or is part of the typing procedure. For the principle of PCR, see the website <http://users.ugent.be/~avierstr/principles/pcr.html>

Fragment-based methods are relatively simple to perform and rather inexpensive. However, the discriminatory power and the reproducibility, especially between laboratories, can vary extensively.

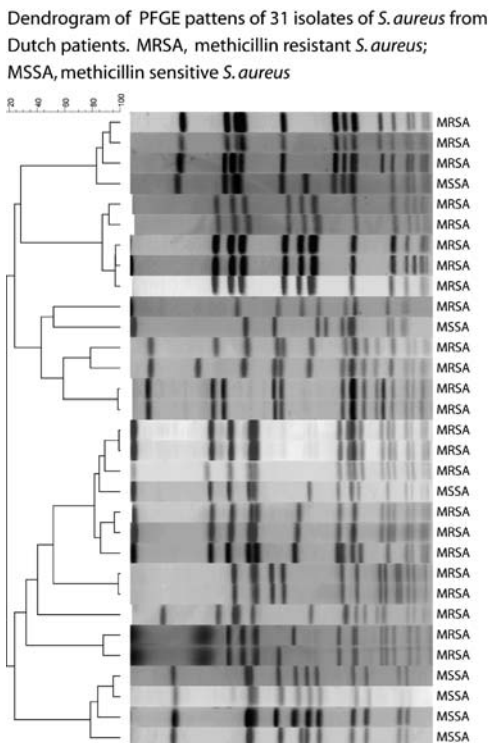
7.3.2 RFLP, PFGE, and Ribotyping

In the last few decades a typing technique called restriction fragment length polymorphism (RFLP) has been used extensively to characterize bacterial isolates. In RFLP, total bacterial DNA is isolated from pure cultures and cleaved with restriction enzymes that recognize short double-stranded sequences. The generated fragments are separated on gels and in the early applications of this technique the patterns were visualized by staining. However, due to the complexity of the bacterial genomes, this technique yields hundreds of bands, too many to use for reliable typing. Therefore, these banding patterns are often simplified by hybridizing the separated fragments with specific DNA probes. A “probe” is an oligonucleotide with a sequence that is known to be complementary to the generated fragment. To accomplish this, the DNA bands are transferred from the agarose gel onto a membrane and used for hybridization. This technique is called Southern blot hybridization or DNA fingerprinting (Snipes et al. 1989). For many bacterial pathogens the ubiquitous but

specific *16S rRNA* genes are used as probes (ribotyping), for example, for MRSA, enterococci, and *Mycobacteria*. The *16S rRNA* gene is usually present in several copies in the bacterial genome. For example, in *E. coli* there are seven copies, and as a result the complex banding pattern of hundreds of bands is reduced to only seven bands, making comparison easier and much more reliable. It is important to realize that in ribotyping and other RFLP-based variations in banding, the profiles are caused by insertions, deletions, and inversions of relatively large segments of DNA and by simple mutations in the restriction enzyme recognition sequence. Often a high degree of polymorphism is associated with repetitive DNA elements. Members of the mycobacterium complex (for example), can be typed very well using insertion sequences (IS) such as the IS6110 element (Arnold 2007; (Reisig et al. 2005; van Embden et al. 1993). The variable number (0–25) of IS6110 copies are dispersed over the mycobacterium genomes and they can be detected by hybridization with specific DNA probes. The IS6110 typing has been proven to be reliable and reproducible within a laboratory if highly standardized protocols for DNA isolation, restriction analysis by gel separation, and hybridization are used (van Embden et al. 1993; Behr and Mostowy 2007). However, exchange of DNA profiles for inter-laboratory comparison is difficult. For example, for tuberculosis outbreaks in Houston, Texas, a US metropolitan area with very high case rates, studies were performed using combined molecular typing with IS6110-based fingerprints, classical epidemiology, and network analysis to reconstruct an outbreak network (Klovdahl et al. 2001). A network is defined as a set of nodes connected together by links of one kind or another. The data were also used to quantify the relative importance of different actors (persons and places), which played a role in the tuberculosis outbreak, showing that the majority of the cases were men who had sex with other anonymous men. These homosexual men were thus proven to be part of a large outbreak and without the IS6110 typing data, their links would have remained uncovered (Klovdahl et al. 2001).

Another widely used approach to reduce the number of bands for RFLP analyses is a typing method called pulsed-field gel electrophoresis (PFGE). In PFGE the extracted whole genome bacterial DNA is digested by restriction enzymes that recognize rare cutting sites (Snell and Wilkins 1986). For example, the restriction enzyme *HindIII* cuts the *Staphylococcus aureus* genome into more than 1,000 fragments (as was used for ribotyping of MRSA). In contrast, the restriction enzyme *SmaI* which is used for PFGE, yields only approximately 25 fragments (Fig. 7.2). These DNA fragments obtained with a rare cutting enzyme are very large (50–100 kb), too large for conventional gel separation. Using special electrophoresis equipment (CHEF, Clamp) with alternating electrical fields, these large fragments can be separated very reproducibly with typical running times of 30–50 hours. PFGE is still considered the gold standard technique for the typing of bacterial strains as it can assist in establishing clonal relationships. However, PFGE is technically demanding and the interchangeability of typing patterns among laboratories is limited because small technical variations in electrophoresis conditions result in different banding patterns.

Fig. 7.2 Example of a PFGE gel



7.3.3 PCR-RFLP

Using PCR-based typing has the advantage that typing can be performed even if very little DNA is available. In some cases where culture of the pathogen is impossible, PCR-RFLP may provide the only method to characterize the agent. However, in most cases it is still necessary to first culture the pathogen to have sufficient pure starting material. In PCR-RFLP, a specific genomic part (a gene or an intergenic region) is amplified and the PCR product is subsequently cleaved with certain restriction enzymes (as in RFLP). The cleaved PCR products are either visualized on gels or separated with an automated sequencer in the fragment analysis mode. The idea is then to screen for the presence of different restriction recognition sites by comparing banding patterns. Only a part of the genome is now used for typing and care should be taken to choose an informative part in order to have a high discriminatory power.

7.3.4 RAPD and AP-PCR

For outbreaks, and particularly for nosocomial outbreaks, it is essential to have typing results at short notice, and for this purpose a “quick and dirty” typing technique

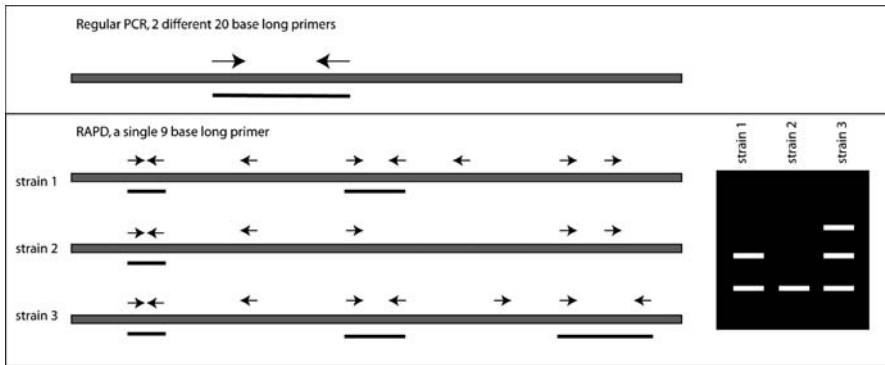


Fig. 7.3 Principle of RAPD. Lines in *below arrows* show PCR products for different strains 1, 2, and 3. The accompanying (*white*) bands on the (*black*) gel are shown for each strain

is available, termed the random amplified polymorphic DNA (RAPD) or the arbitrarily primed PCR (AP-PCR) (Williams et al. 1990). RAPD requires the availability of pure DNA from cultures, however, and the isolation and purification of DNA is a step that limits the speed of typing. Short nonspecific oligonucleotides (8–12 nucleotides) are used to serve both as forward and as reverse primer to generate random PCR products. Informative AP-PCR typically yields a number of PCR products ranging from 300 to 2500 base pairs that can be visualized on agarose gels (Fig. 7.3). In a multicenter study by van Belkum et al. (1995), the AP-PCR typing performance was assessed with respect to reproducibility and discriminatory power using 60 well-defined *Staphylococcus* strains (van Belkum et al. 1995). These *S. aureus* strains were collected at five documented outbreaks and also included sporadic cases. The DNA samples were centrally isolated and distributed to seven laboratories. Each laboratory used the same set of three random primers for AP typing but still, even in these controlled circumstances, the inter-laboratory variation was extensive. Although the outbreaks could be recognized in the resulting clusters as also established by PFGE, the general conclusion is that AP-PCR should not be used for *S. aureus* outbreak analysis if performed in different laboratories (Deplano et al. 2000; van Belkum et al. 1995). In general, the AP-PCR method is suited only for comparison of small numbers of isolates that are processed simultaneously. It is not suitable for longitudinal comparison, not even within the same laboratory (van Belkum et al. 2007).

7.3.5 AFLP

The benefits of RFLP and the sensitivity of PCR can be combined as was devised in amplified fragment length polymorphism (AFLP) (Vos et al. 1995). This high-resolution genomic fingerprinting method starts with purified pathogen DNA that is cleaved with two specific restriction enzymes. One of the two restriction enzymes,

e.g., *MseI*, yields many small fragments and the other, e.g., *EcoRI*, yields relatively few, larger fragments (Koeleman et al. 1998). After the cleavage with the restriction enzymes, two types of synthetic oligonucleotides (linkers), one that fits at the *MseI* cut end and the other that fits only at the *EcoRI* end, are ligated (“colated”) onto the PCR fragments. Subsequently, PCR amplification with two different primers that hybridize either of the two linkers is performed and the primer that anneals with the *EcoRI* linker carries a 5'-fluorescent label. Using AFLP, only those small fragments that are cut by *MseI* at one end and by *EcoRI* at the other end are amplified and visualized. The resulting PCR products ranging from 50 to 500 base pairs are separated on an automated DNA sequencer. It is also possible to use “selective” primers; these primers contain one extra nucleotide at their 3'-end which limits the number of PCR fragments formed, thereby further reducing the complexity of the profile. AFLP was also used for typing members of the *Mycobacterium tuberculosis* complex and many polymorphic markers for the different subspecies were identified (van den Braak et al. 2004). Although AFLP yields a high-resolution typing technique, interpretation of profiles is notoriously difficult due to the variation in the intensity of bands. This problem is caused by the variation in ligation and PCR. As a result, inter-laboratory exchange of AFLP data is virtually impossible.

Comparison of typing performance of AP-PCR (= RAPD), PFGE, and AFLP was described in a study of nosocomial outbreaks of Legionnaires disease in Germany (Jonas et al. 2000). *Legionella pneumophila* strains were obtained from both patient bronchoalveolar lavages and environmental sources. All three typing procedures found one predominant genotype that was associated with the hospital outbreaks. The discriminatory power differed however, with AP-PCR being the least and AFLP the most discriminating method. Because of this, and its simplicity and reproducibility, Jonas et al. concluded that AFLP was the most effective typing technique (Jonas et al. 2000). Of course this is open to debate and always depends on the epidemiological questions that need to be answered.

7.3.6 MLVA

The last fragment-based typing technique described here is called MLVA and is based on the variation in the number of tandem repeats. Tandem repeat loci are made up of two or more identical or nearly identical short DNA sequences that are not interspersed with any intervening DNA sequence. Tandem repeats are the result of errors of the DNA polymerase, which incorrectly copies these segments by a mechanism called slipped strand mispairing. The DNA polymerase stumbles in certain regions in the genome and as a result some DNA regions are duplicated or deleted. This malfunction of the DNA polymerase may happen several times and will cause some regions to be multiplied several times and the size of these tandem repeat units may range from 3 to 100 bp. These stutter regions result in variation in the number of repeats; hence their name “variable number of tandem repeats” (VNTR) loci (van Belkum et al. 1998). See also Fig. 7.4.

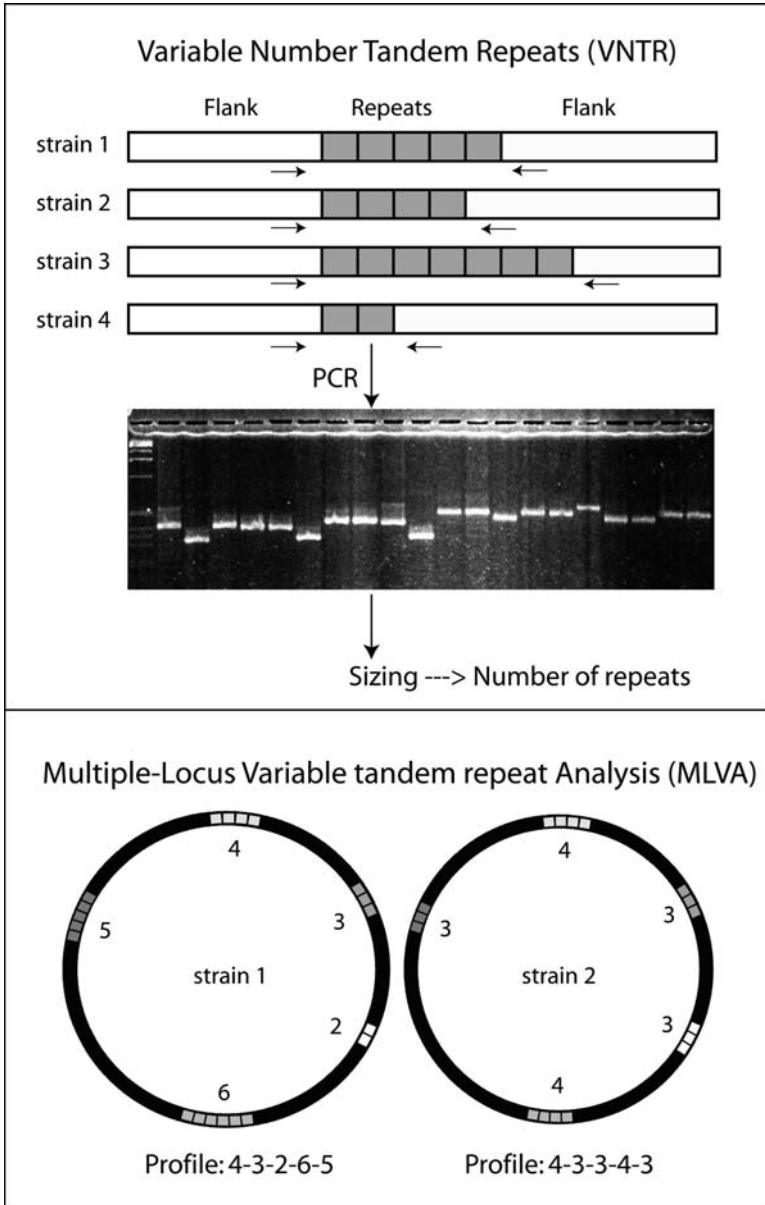


Fig. 7.4 Variable number of tandem repeats (VNTRs). VNTRs are polymorphic DNA sequences composed of different numbers of a repeated “core” sequence arranged sequentially. The size of the core sequence can vary from 8 to 100 bp in different VNTRs, and the number of repeats present at a VNTR locus also varies widely. Although many different VNTRs have been identified in genomes of both human and pathogen origin, their function is not currently known

The polymorphism in repeat loci is utilized in a typing method called multiple-locus VNTR analysis (MLVA) (Fig. 7.4). MLVA is a method that utilizes variation in 4–15 regions of tandemly repeated DNA in the genome (van Belkum et al. 2007). By performing PCRs spanning these repeats, the size of the product can be assessed by gel electrophoresis or by sequencing analysis. From the PCR product size, the number of repeats can be deduced. By combining the number of repeats of several repeat loci, a multidigit, specific strain code is obtained and these profiles can be used for clustering analysis. The use of automated DNA sequencers to perform the sizing of the PCR product makes MLVA a very reliable method and the numerical nature of the data makes it suitable for inter-laboratory exchange. The first published study on the use of MLVA was in 1993 and in this study, van Belkum et al. used the variation in various VNTR loci of *Haemophilus influenzae* (van Belkum et al. 1994). For *Mycobacteria*, MLVA (MIRU, mycobacterium interspersed repeat units-VNTR) was successfully used to type members of the tuberculosis complex (van Deutekom et al. 2005; Behr and Mostowy 2007). Given its technical simplicity, MLVA may have a successful future as it performed well compared to other genotyping methods for a variety of bacterial species so far (Schouls et al. 2006; Schouls et al. 2004; Lindstedt 2005; Francois et al. 2005; Noller et al. 2006).

Many fragment-based typing techniques are available today and for some purposes it may be needed to combine typing techniques to provide answers. For example, IS 6110 typing for TB characterization may not be discriminatory if less than five IS6110 copies are present. In TB typing studies, combinations of IS6110, spoligotyping (see below), MIRU-VNTR, and/or AFLP were used (Bauer et al. 1999; van Deutekom et al. 2005; Behr and Mostowy 2007).

7.4 Hybridization Arrays

7.4.1 Spoligotyping

Spoligotyping is an acronym for *spacer oligo typing* and is a variation of the reverse line blot hybridization technique (Kamerbeek et al. 1997). In reverse line blotting, oligonucleotide probes are bound covalently to a membrane in parallel lines and hybridized with biotin-labeled PCR products in parallel lines, perpendicular to the oligonucleotide lines (Fig. 7.5). The intersections of the lines, where PCR products hybridize, are visualized by enhanced chemiluminescence. This generates easy to read fingerprints that are comparable to barcode patterns. For example, spoligotyping is used as a technique that analyzes a genomic region present in *M. tuberculosis* where virtually identical 36-bp direct repeats are interspersed with unique 35–40-bp spacer regions. Primers that recognize the direct repeats are used to amplify the direct repeat region including the spacer sequences between these variable repeats (Fig. 7.5). The resulting mixture of PCR products is used in a reverse line blot hybridization using immobilized spacer-specific oligonucleotide probes (Kamerbeek et al. 1997). Because some MTC members such as *M. leprae* are not cultivable, spoligotyping is of general use for all MTC members. The

SpoligoTyping

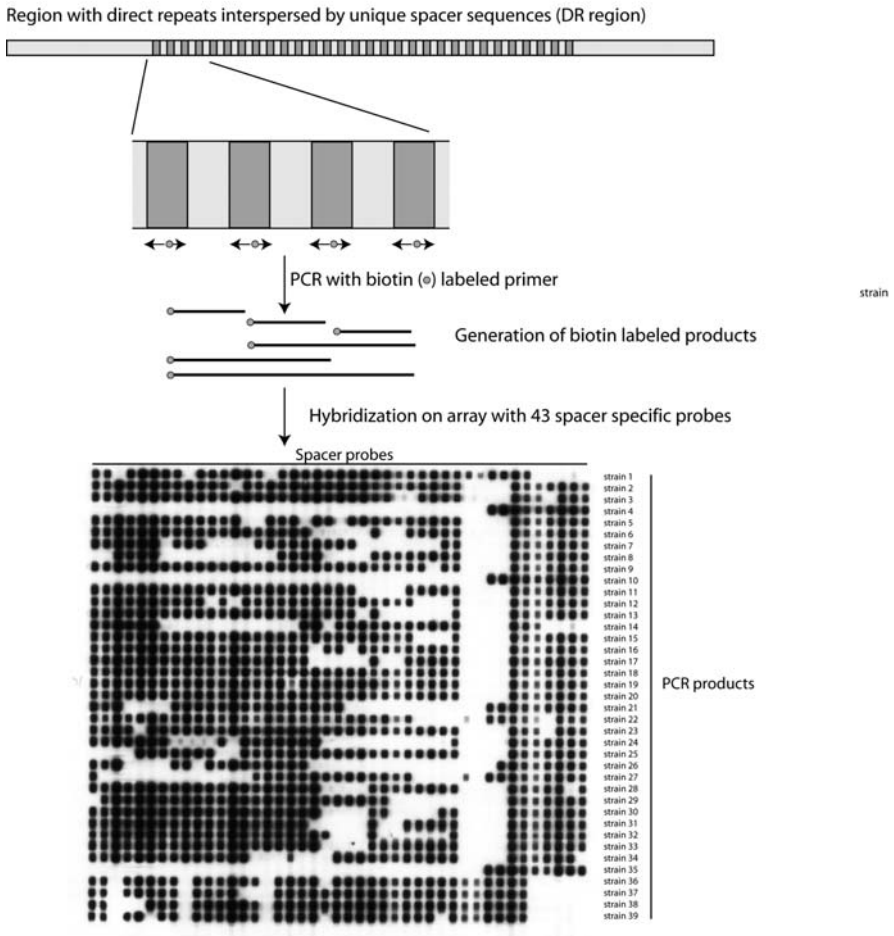


Fig. 7.5 SpoligoTyping. Spacer oligonucleotide typing (spoligoTyping) is a molecular method used to differentiate, for example, *M. tuberculosis complex* isolates. This method is based on the analysis of polymorphisms in the *M. tuberculosis complex* direct repeat (DR) chromosomal region consisting of identical 36-bp DRs alternating with 35-to 41-bp unique spacer sequences. The method is PCR based and is hence more rapid and easier to perform than the standard typing technique based on IS6110 profiling. SpoligoTyping can also be performed directly from *M. tuberculosis* organisms, even those that are nonviable or that are found in tissues in paraffin-embedded blocks or in archeological samples

discriminative power of spoligoTyping for typing MTC is lower than that of IS6110 fingerprinting (Kremer et al. 2005; van der Zanden et al. 2002). However, Bauer et al. found spoligoTyping to be very useful to discriminate low copy number IS6110 MTC strains in Denmark, where tuberculosis caused by this type of strain increased in the last decade, supposedly due to influx of immigrants with tuberculosis (Bauer et al. 1999).

7.4.2 Microarrays

Microarray analysis can be used to perform full-length typing of genomes, even of large bacterial or parasitic genomes. Whole genome arrays detect the presence or the absence of similar DNA regions in sufficiently related microorganisms, allowing genome-wide comparison of their genetic contents (Garaizar et al. 2006). In this technique, called comparative genomic hybridization (CGH), fluorescently labeled genomic DNA of a microorganism is hybridized with a very large collection of DNA probes fixed on a solid support. The probes can be PCR products or synthetic oligonucleotide probes that are usually about 70 base pairs in length and are mostly specific for all the open reading frames (ORFs or genes) present in the bacterial genome of interest. The probes, which are designed based on the strain from which a genome sequence is available, can be deposited on support materials such as nitrocellulose membranes, but mostly glass slides are used. The latter support material permits the deposition of up to 10^4 – 10^5 probes. Differences between strains are expressed as the presence or the absence of genes that are present in the indicator strain, i.e., the strain from which the probe sequences were derived. If short oligonucleotide probes are used, it is also possible to detect single nucleotide polymorphisms (SNPs), for which probes covering over 10^5 polymorphisms can be used in a single hybridization, permitting excellent discrimination between different strains of a species (Garaizar et al. 2006). The major drawback of CGH is that only already known genes are tested. New genes, not present in the strain used to design the microarray, will never be detected. The costs for the microarray technology are relatively high, for the time being too expensive for routine typing.

7.5 Sequence-Based Typing (SBT) Analysis

In recent years, DNA sequence-based typing has started to replace the fragment-based typing methods because SBT has fewer problems with reproducibility and portability. DNA sequence data have the major advantage that they are unambiguous and independent of the method used to obtain the DNA sequence. This has resulted in portable methods that yield robust data that are highly suitable for inter-laboratory exchange.

Viral genomes are rather small compared to bacterial genomes. For example, the hepatitis B virus (HBV) genome is only around 3,200 base pairs in size and the hepatitis A virus (HAV) is 7,500 base pairs in size. Typing viral pathogens is often performed by sequencing parts of the genome (either coding/genes or non-coding). The sequenced regions need to be polymorphic to discriminate within a species. Choosing regions that are conserved will not give enough discriminatory power and regions that are highly variable will lose the link in the dissemination chain, making them unsuitable for epidemiological purposes. The discriminatory power has to be assessed by first validating the chosen region. Suitable genetic regions are, for example, genes or parts thereof that encode surface structures, such as the S-gene in HBV and the VP genes in HAV. Variability in genes encoding surface

structures is mostly caused by the selective immune pressure of the host. Too much variability would not be suitable for clustering analysis, with every infected host having its own strain of virus. On the other hand, choosing genomic regions that are too conserved in sequence will yield very large clusters, linking persons who never were in contact. To find out which level of variation suits the epidemiological questions best, it is absolutely required to validate the typing method by linking it to sound classical epidemiological information.

7.5.1 SLST

This “single-locus sequence typing” (SLST) was found useful in molecular epidemiological studies of hepatitis A, B, and C viruses in our laboratory and some examples are given below in the paragraph dealing with cluster analysis. For hepatitis B virus, an international database HepSEQ is accessible at <http://www.hpa-bioinfodatabases.org.uk/hepatitis/main.php>

In recent years, single-locus sequence typing has also been used to type bacterial pathogens. One of the currently best known pathogens for which SLST has been utilized is *S. aureus* and in particular methicillin-resistant *S. aureus* (MRSA). In *S. aureus* the gene encoding the immunoglobulin G-binding protein A contains a tandem repeat region that is variable both in number of repeats and in the DNA sequence of these repeats. In the so-called Spa-(*Staphylococcal protein A*) sequence typing, a PCR product encompassing the repeat region is sequenced and the sequence is used to create profiles representing the composition of the repeat region (Harmsen et al. 2003). This results in the assignment of Spa types, each of which has its own Spa profile. As an example, Spa type t073 has an Spa profile of r08r16r02r16r13r17r34r16r34. The profile shows that this Spa type contains nine tandem repeats and in addition the composition of the specific repeats. In the given profile, for instance, r08 represents the sequence **GAGGAAGACAACAACAAGCCTGGT** and r16 is similar to this but differs at the positions in bold **AAAGAAGACGGCAACAAACCTGGT**. Currently more than 200 different repeat sequences and nearly 3,500 Spa types have been identified in the *S. aureus* genomes.

7.5.2 MLST

For bacteria the best known sequence-based typing is *multi*locus sequence typing (MLST) in which the DNA sequence of a set (mostly seven) of housekeeping genes is determined by sequencing polymorphic 400–500 base pair segments of each of these genes. Housekeeping genes are encoding components that are essential for the bacterium (Maiden et al. 1998). As a result these genes are always present and change only slowly over time. In MLST each of the resulting housekeeping gene sequences is assigned an allele number. Allele numbers are consecutive numbers that have been assigned to the sequence variants of each housekeeping gene and

thus do not reflect the degree of similarity between alleles. Each strain can be characterized as a sequence type (ST), representing an allelic profile, which is an ordered string of allele numbers separated by dashes, e.g., ST-100 in *Neisseria meningitidis* represents allelic profile 6-6-9-1-26-36-32. MLST is a portable, unambiguous, and highly discriminating genotyping system that can be used for many bacterial species and even for some other microorganisms such as *Candida albicans*. However, some bacterial species, e.g., *M. tuberculosis*, are too clonal, rendering MLST useless as a typing technique. Although the availability of pure cultures is required to perform MLST, isolation and purification of DNA are not. The major disadvantages of the MLST are its labor-intensive nature and the high costs associated with sequencing. In order to characterize a single isolate, one needs to perform seven PCRs and subsequently 14 sequence analyses (two directions per PCR product). The MLST approach has been successful for the unambiguous characterization of isolates of many bacterial species and other microorganisms and a number of libraries can be accessed over the Internet (www.mlst.net, website visited November 2007).

7.5.3 Whole Genome Sequencing

MLST allows clustering of isolates using an unambiguous DNA sequence and has been a major step forward in genotyping methodology. However, seven gene sequences obtained with MLST for *L. pneumophila* represent only 0.1% of the total genome. Increasingly the full-length DNA sequences of organisms, particularly of bacterial species, are being determined. Many of these sequences are available in the public domain, e.g., in GenBank:

<http://www.ncbi.nlm.nih.gov/sites/entrez>

At the time of the writing of this chapter, 634 microbial genomes have been completely sequenced and are available online

<http://www.ncbi.nlm.nih.gov/sites/entrez> and

<http://www.integratedgenomics.com>

Sequencing of another 965 microbial genomes is in progress and the number will rapidly expand due to the fact that new high-throughput sequencing methods have become available. It is now possible to sequence a complete bacterial genome within a week's time. Comparison of complete genome sequences is the ultimate genotyping method. However, costs of sequencing a complete bacterial genome are still too high to use this ultimate form of genotyping.

7.5.4 SNP Genotyping

Single nucleotide polymorphisms (SNPs) can be used successfully for genome-wide typing. It involves the determination of the nucleotide base that is known to be variable at a defined position in the genome in a given isolate. It is a more efficient typing method than full genome sequencing not only because of lower costs to perform but also because the results can be easily shared by different users if standard sets of SNPs per pathogen are analyzed (van Belkum et al. 2007). Variable positions

useful for typing must have been discovered prior to the application of SNP genotyping. Once a set of SNPs has been selected, different SNP genotyping methods such as Sanger sequencing or pyrosequencing can be applied. Newer methods are being developed at a fast pace, for example, mini-sequencing and Luminex technology (van Belkum et al. 2007; Syvanen 2005; Dunbar 2006). Mini-sequencing uses primers and a mixture of all four dideoxynucleotides, each with their specific fluorescent label. Each primer borders an SNP region and is extended by only a single base, and the incorporated base can be discriminated by fluorescence after capillary electrophoresis. In pathogens with highly homologous genomes, such as *E. coli* O157:H7, *M. tuberculosis*, *Salmonella enterica* serotype Typhi, and *Bacillus anthracis*, SNP genotyping methods can be very helpful to quickly discriminate them (van Belkum et al. 2007).

7.6 Cluster Analysis

Epidemic histories can be assessed by constructing phylogenetic trees and subsequent cluster analysis. In a phylogenetic tree, or dendrogram, groups can be assigned by distinguishing clusters defined by a distance measure, such as the percentage nucleotide difference (which is a crude distance measure). These clusters are then, by consensus, defined as specific genotypes. For a conserved sequence region (i.e., for HAV), a genetic distance of >15% may indicate different genotypes, but genetic differences between 15 and 5% indicate different “sub-genotypes” (in case of HAV: 1A and 1B). So, clusters may represent either (sub)-genotypes or even groups of strains with less genetic distance between them.

In general, to use typing information for answering epidemiological questions, information collected in libraries needs to be structured by performing cluster analysis. To determine which of the ensuing strains are found in a particular outbreak, or a specific high-risk group, it is necessary to combine clinical and conventional epidemiological information with the typing data. To this end, different clustering of the genetic typing data must be performed based on the variation of genetic distance measures. Then the resulting clustering structure has to be compared to epidemiological data and analyzed for corresponding structural properties. For example, if a very fine genetic distance measure is used to define clustering, only small clusters will be found. These small genetic differences would have arisen by chance, as mutations occur at all times in all genomes of living beings, but they may also reflect pressures of the host immune response, generating polymorphism between source and contact strains. However, a small genetic distance that reflects only one or two mutations may point to the presence of the same epidemic strain. A clustering based on a coarser distance measure might be able to bring larger epidemiological structures to the foreground, such as transmission clusters that connect cases within a specific risk group.

Algorithms are needed to construct clusters starting from an available typing library. The BURST algorithm (<http://eburst.mlst.net/>) first identifies mutually exclusive groups of related genotypes in the population (typically an MLST

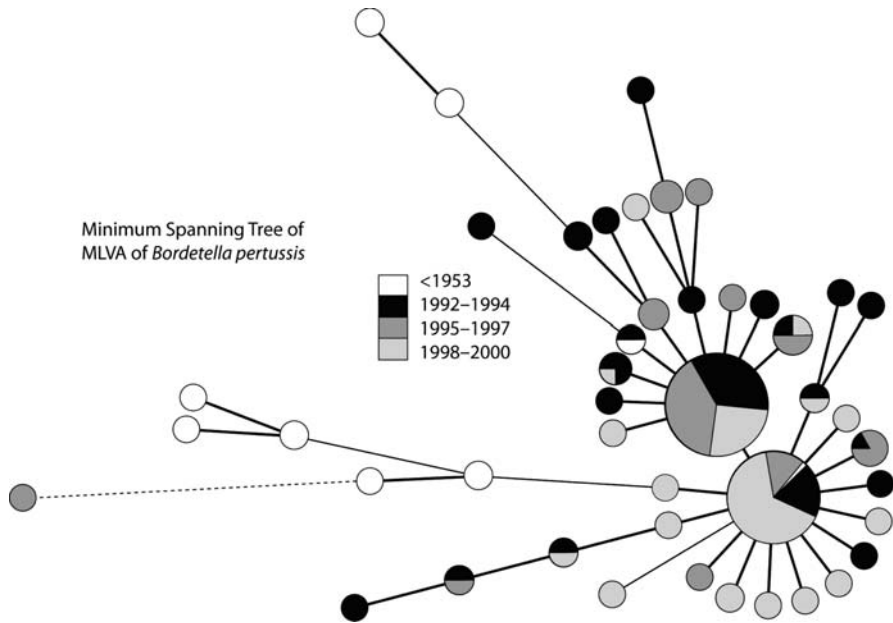


Fig. 7.6 Minimum spanning tree. Example for *B. pertussis*, the causative agent of whooping cough

database) and attempts to identify the founding genotype (sequence type or ST) of each group. The algorithm then predicts the descent from the founding genotype to the other genotypes in the group, displaying the output as a radial diagram, centered on the predicted founding genotype. The presentation of such a grouping is called a minimum spanning tree (MST) and software packages are available to construct these MSTs. In Fig. 7.6 the size of the circles corresponds to the amount of strains; colors may indicate certain properties of strains or their hosts.

The MSTs may depict either single-locus variations (the profile varies at one locus) or dual-locus variations (the profile varies at two loci), by connecting the circles. The MST procedure is based on an old algorithm that has been adapted for clustering character-based data such as MLST and MLVA.

7.6.1 Cluster Analysis for Nosocomial and Community-Acquired Outbreaks

Community-acquired outbreaks can be studied using previously built databases that contain typing data connected to epidemiological data. A database such as PulseNet for foodborne infections proved very successful in aiding source tracing and pointing out reservoirs of infection. PulseNet participants perform standardized molecular subtyping by pulsed-field gel electrophoresis to distinguish foodborne

disease-causing bacteria such as *E. coli* O157:H7, *Salmonella*, *Shigella*, *Listeria*, or *Campylobacter* (website: <http://www.cdc.gov/pulsenet/>). The PulseNet network initially started in 1996 in the United States, but nowadays these PFGE-based databases have been developed and made accessible also in Canada, Latin America, Japan, and Europe (<http://www.pulsenet-europe.org/>) and in other parts of the world.

Viral foodborne infections by noroviruses, which cannot be cultured, are widespread. An example of the additional use of typing is given in Koek et al. (Koek et al. 2006). In this study, we describe the molecular epidemiology of a group of nine outbreaks associated with a catering firm and two outbreaks, 5 months apart, in a hospital in Amsterdam, the Netherlands. All outbreaks were typed to confirm their linkage, and the hospital-related cases were studied to see if the two outbreaks were caused by one persisting norovirus strain or by a reintroduction after 5 months. For the outbreaks associated with the catering firm, one norovirus genogroup I strain was found which was identical in sequence among customers and employees of the caterer. This was not the strain that predominantly circulated in 2002/2003 in and around Amsterdam, which was the norovirus genogroup II4 “new variant” (GgII4nv) strain. In the Amsterdam hospital, the two outbreaks were caused by this predominant GgII4nv type, and we argue that NV was most likely reintroduced in the second outbreak from the Amsterdam community.

7.6.2 Cluster Analysis and Linkage to Risk Groups

From DNA sequences it is possible to create phylogenetic trees and to determine the reliability of the tree branchings by bootstrapping. Bootstrap values >90% indicate that a (sub)cluster can be discerned with a high probability. This way of typing and clustering is mostly used for viruses with rather small genomes.

7.6.2.1 Human Immunodeficiency Virus

The structure of sexual contact networks can be reconstructed from interview data and in some cases this provided valuable insights into the spread of the infection. For HIV-1 however, the long period of infectivity and the anonymous sexual contacts made the interpretation very difficult, producing discrepancies in the networks. Using viral genotype data from large sets of HIV-1 *pol* and *gag* sequences, which were initially collected to monitor therapy resistance to antivirals, enabled Lewis and colleagues recently to derive the network structure of HIV-1 transmission among homosexual men in London (Lewis et al. 2008; Pilcher et al. 2008). Nine large clusters were discerned on the basis of genetic distance. Dated phylogenies with a molecular clock-like calculation indicated that 65% of the HIV-1 transmissions took place between 1995 and 2000 and that 25% occurred within 6 months after infection. The quantitative description, also called “phylogenetics” (Grenfell et al. 2004), is important for parametrization of epidemiological models and in designing intervention strategies (Lewis et al. 2008).

Many HIV phylogeny studies have been published and others are still in progress. A recent Chinese study is informative for the use of phylogenies in typing sequences from HIV-1-infected drug users in the Yunnan Province, which borders Myanmar and Tibet, countries known to be involved in illegal drug transports (Zhang et al. 2006). Recombinant circulating HIV-1 strains of types BC and AE were found to circulate codominantly among the 321 HIV-1-infected and analyzed persons. The type BC strain was strongly associated with intravenous drug behavior, whereas the type AE strain was mainly sexually transmitted. This last type AE appears to be on the rise and forms a threat to the general population. Aids education and prevention efforts in the general population are therefore urgently needed (Zhang et al. 2006).

7.6.2.2 Hepatitis A Virus

Clustering analysis in hepatitis A epidemiology is given in Tjon et al. 2007. In Amsterdam, the Netherlands, the patterns of introduction and transmission of hepatitis A virus (HAV) were investigated from 2001 to 2004 and HAV strains were divided according to two risk groups: (1) travelers and their contacts, who were most often infected with HAV subgenotype 1B strains, and (2) homosexual men and their contacts, who were shown to have subgenotype 1A strains. Among travelers many sporadic cases were found, and the clusters were small and limited in time but introduced frequently into the population, mostly in the second half of each calendar year, indicating a seasonal pattern of introduction and transmission after the summer holidays. These introductions were especially by Dutch children from parents originating from hepatitis A-endemic countries, like Morocco. Among men who have sex with men (MSM), the clusters were bigger and remained present for a longer time; sporadic cases were few, and introduction of new strains occurred only occasionally but throughout the year. Our findings indicate that travelers frequently import new HAV strains into Amsterdam, but they are limited in the extent and season of their spread. In contrast, HAV is endemic among the male homosexual and bisexual population, and the same strain spreads to many individuals without a seasonal pattern.

Large outbreaks of hepatitis A have also occurred elsewhere in Europe affecting MSM in countries such as Denmark, Germany, Norway, Spain, Sweden, and the United Kingdom during the period 1997–2005. An international collaboration was formed between these countries to determine if the strains involved in these outbreaks were genetically related. Part of the genetic regions coding for HAV capsid genes were sequenced and compared. The majority of the HAV strains found among homosexual men from different European countries formed a closely related cluster, named MSM1, belonging to genotype IA. Different HAV strains circulated among other risk groups in these countries during the same period, indicating that specific strains were circulating among MSM exclusively. Similar strains found among homosexual men from 1997 to 2005 indicate that these HAV strains have been circulating within this group for a long time. The homosexual communities across Europe are probably large enough to sustain continued circulation of HAV strains for years, resulting in an endemic situation among MSM (Stene-Johansen et al. 2007).

7.6.2.3 *Neisseria gonorrhoeae*

Bacterial typing also needs validation of cluster formation. In London and Sheffield, UK, sexual links were described and compared among people with gonorrhea, caused by the sexually transmitted *N. gonorrhoeae*. Most cases concerned homosexual men and fewer female sex workers. In Sheffield, large, linked heterosexual networks identified were associated with local contacts but the networks in London were more difficult to trace back, due to anonymous contacts (Day et al. 1998). Subsequently, the use of gonococcal *opa* typing, based on the polymorphisms of the *opa* genes, suggested a highly connected population in Sheffield where almost 80% of cases had shared profiles. In London the *opa* typing could also link infections that would otherwise have remained unlinked, and typing may thus aid interventions to control endemic gonorrhea (Ward et al. 2000). In Amsterdam, the Netherlands, *N. gonorrhoeae* strains were also collected for several years and patients were subjected to a questionnaire pertaining to sexual risk behavior and sexual partners in the 6 months prior to the diagnosis (Kolader et al. 2006). The *N. gonorrhoeae* isolates were all genotyped using PCR–RFLP of the *porin* and *opa* genes. There were 11 clusters of ≥ 20 patients; in seven clusters, almost all patients were MSM, three clusters contained mainly heterosexual men and women, and one cluster was formed by equal proportions of MSM and heterosexual male and female patients. However, the various clusters also differed in characteristics such as types of coinfections, numbers of sexual partners, Internet use to seek sexual partners, and locations of sexual encounters. Molecular epidemiology by typing of gonococcal isolates thus revealed core groups and clusters of MSM and heterosexual patients that probably indicate distinct transmission networks (Kolader et al. 2006).

7.6.2.4 Methicillin-Resistant *S. aureus*

An example of the additional use of clusters containing different sequence types (STs) as defined by MLST is given here for methicillin-resistant *S. aureus* (MRSA). MRSA originated through the transfer of the mobile resistant determinant staphylococcal cassette chromosome *mec* (SSC*mec*) into sensitive *S. aureus* isolates. Community-associated (CA) MRSA have been reported all over the world. The prevalence of MRSA is less than 1% in the Netherlands because of the “search and destroy” policy in hospitals affected with MRSA. In the case where two indiscriminate isolates are found, costly measures are taken such as very strict hygiene, cohorting of patients and staff in hospitals, closing of wards, and postponing or rescheduling surgery. All those in contact with the index patients are screened for MRSA carriage. Quick typing results are required to save time and money to keep the Dutch prevalence low. The MRSA prevalence in hospitals is much higher, however, in surrounding countries such as the United Kingdom, France, and south European countries, with prevalences up to 30%. Since 2003 a new clone of CA MRSA has been found in farm animals (pigs and veal calves) and in humans in direct contact with these animals (Voss et al. 2005). The first recognized case with pig MRSA was a six-month-old girl who was found MRSA positive in a Dutch

hospital. She was the daughter of a pig farmer (Huijsdens et al. 2006). In a still ongoing outbreak in the Netherlands and Belgium, this MRSA strain was found to be “untypeable” by PFGE, which was considered to be the gold standard typing procedure. PFGE failed to give adequate answers because the *SmaI* enzyme could not cleave the DNA derived from outbreak-associated SA strains. Spa (Staphylococcus protein A) typing (see above) and MLST showed that strains of this MRSA clone belong to a number of closely related Spa types, all of which correspond to MLST ST-398 (Voss et al. 2005).

7.7 Limitations of Typing

The first limitation of molecular typing is that the specific nucleic acid content (usually the genome) of the pathogen is not always available for analysis. For example, in the case of whooping cough, the patient shows prolonged respiratory symptoms due to the toxins that were excreted by the etiological agent, the bacterium *Bordetella pertussis*. However, these symptoms often become apparent only after the pathogen itself was cleared by the time samples are taken, resulting in negative culture. People may carry the pathogen, but due to clearance or (partial) immunity the pathogen is present in such a low load that isolation does not succeed even though transmission may still occur.

There are several assumptions concerning the use of molecular typing techniques for the interpretation of outbreak data (Singh et al. 2006):

1. Isolates associated with an outbreak are of recent progeny of a single precursor or clone
2. All outbreak-associated descendants have the same genotype
3. Epidemiologically unrelated isolates have different genotypes

For frequently used techniques such as AP-PCR, PFGE and also AFLP, the reproducibility suffers inevitably because the electromobility of the DNA fragments varies between lanes in gels and between runs on separate days. But the variation between laboratories is even more extensive due to differences in DNA isolation procedures and the use of different gel instrumentation.

The major disadvantages of PFGE are the need for specific equipment, the complexity of performing the assay, and above all, the lengthy time to complete the results (van Deutekom et al. 2005). Nevertheless, the discriminatory power of PFGE is high, so replacing this technique will require at least the same index of discrimination. Also MLST clearly has its limitations. If housekeeping genes that do not show enough polymorphism are chosen, for example, in homomorphic, recently introduced pathogens, it may not be possible to separate outbreaks from other, sporadic, cases. Similarly, VNTR loci used for MLVA may be invariable or display hypervariability. The rate of change in the number of repeats may differ between the VNTR loci, influencing the stability of an MLVA pattern. For example, for *E. coli* O157:H7

(the hamburger bacterium), the TR2 locus showed a much higher degree of variation than the other six loci tested (Noller et al. 2006).

The microarrays are presently only sporadically used for typing because of the comparatively low availability of complete genome arrays and the high cost per reaction (Garaizar et al. 2006). Another limitation is that the genetic reservoir may be unstable. Mutations could arise even in clonal strains being induced during culture due to longer duration or temperature variation. Conventional assays such as PFGE or IS typing will not detect this because their discriminatory power is much less compared to full genome microarray analysis.

Typing guidelines have been described to decide when strains or isolates are indistinguishable (Tenover 1995; van Belkum 2007). It is clear however that these rules pertain to the typing method that was chosen because the underlying biological phenomena are not the same. In case of fragment-based typing, the absolute number of band differences or percentage similarity is often taken into account to define relatedness. The absolute number may differ with the possibility that more mutational events lead to less banding differences. The percentage similarity is independent of banding patterns but it is clearly influenced by the level of tolerance in band position. Therefore reference samples should be included in each analysis for normalization (for example, for PFGE). Sequence-based typing methods do not need such reference samples but the quality of the sequences has to be thoroughly checked (van Belkum et al. 2007).

It is important to realize that the genetic markers used for typing usually do not reflect the virulence of isolates. Some markers may be mere coincidental indicators for hypervirulent genotypes or genogroups. The results obtained by genotyping are often used as a starting point for further investigation into the pathogenic properties of subgroups of microorganisms.

7.8 Concluding Remarks

New molecular typing methods have substantially helped to better understand epidemiological observations. It must be emphasized however that typing results never stand alone and need to be interpreted in the context of all available demographical, clinical, and other epidemiological data (van Belkum et al. 2007). In general, the sequence or character-based typing techniques such as MLST and MLVA are superior in that their data can be documented unambiguously, allowing direct comparison of results between laboratories, and it is expected that these methods will prevail in the near future. The PCR-based nature of MLVA gives it a speed that is presently not matched by other typing techniques with equal reproducibility. The few handling steps make it amenable for complete automation and adaptable to future fragment separation techniques (Lindstedt 2005). The typing data are stored in publicly accessible databases or libraries on the Internet, which increasingly also contain epidemiologically relevant information. Building such databases relies on the standardization or at least the harmonization of typing methods (Harmony

website: <http://harmony-microbe.net/index.htm>). A clean working hypothesis must be formulated to guide the choice of the typing method, since different questions may require answers with different levels of discriminatory power. Translating typing results into clinical practice is a very important endpoint of a typing exercise. Real-time typing may now be feasible, but next to speedy typing results, the feedback to all parties involved (in an outbreak) is of major importance. It is also essential that it is clear-cut what appropriate action is needed once indistinguishable isolates are encountered, as happens in the case of an outbreak. Thus, multidisciplinary work provides the basis for outbreak investigation, disease control, and pathogen surveillance.

Acknowledgment The authors wish to thank Prof. Dr. R.A. Coutinho for his valuable comments to improve the readability for non-laboratory trained persons.

Definitions

Adapted from (Tenover et al. 1995; van Belkum et al. 2007)

Isolate: An isolate is a pure culture of bacteria or virus from a primary patient sample. In the molecular era it may however also be a DNA sample isolated from the infected site containing the full genomic content of the pathogen.

Strain: A strain is an isolate or a group of isolates that can be distinguished from other isolates of the same genus and species by phenotypic or genotypic characteristics. A strain is a descriptive subdivision of a species.

Epidemiologically related strains: These isolates are cultured from specimens collected from patients or their excretions (feces, fomites), or from the environment during a discrete time frame, or from a well-defined (geographical) area as part of an epidemiologic investigation that suggests that the isolates may be derived from a common source.

Genetically related strains (clones): These isolates are (almost) indistinguishable from each other by one or a variety of genetic tests, supporting the suggestion that they are derived from a common ancestor.

Cluster or clonal complex
Clonal complex definition: The term cluster is used to indicate isolates with identical or highly similar DNA typing results (fingerprints) but also the group of persons (patients) from whom these isolates were derived.

Outbreak strain: Outbreak strains are isolates of the same species that are epidemiologically related *and* genetically related. Such isolates are presumed to be clonally related since they have common genotypes and phenotypes.

Epidemic strain: Isolates that are frequently recovered from infected patients in a particular health-care setting or community and that are genetically closely related, but for which no direct or epidemiologic relation can be established. Their common origin may be more temporally distant from those of outbreak strains.

References

- Arnold C. (2007) Molecular evolution of *Mycobacterium tuberculosis*. *Clin Microbiol Infect* 13: 120–128
- Bauer J, Andersen AB, Kremer K, and Miorner H (1999) Usefulness of spoligotyping To discriminate IS6110 low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. *J Clin Microbiol* 37: 2602–2606
- Behr MA, and Mostowy S (2007) Molecular tools for typing and branding the tubercle bacillus. *Curr Mol Med* 7: 309–317
- Blanc DS. (2004) The use of molecular typing for epidemiological surveillance and investigation of endemic nosocomial infections. *Infect Genet Evol* 4: 193–197
- Day S, Ward H, Ghani A, Bell G, Goan U, Parker M et al. (1998) Sexual histories, partnerships and networks associated with the transmission of gonorrhoea. *Int J STD AIDS* 9: 666–671
- Deplano A, Schuermans A, Van Eldere J, Witte W, Meugnier H, Etienne J et al. (2000) Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* strains by repetitive-element PCR analysis. The European Study Group on Epidemiological Markers of the ESCMID. *J Clin Microbiol* 38: 3527–3533
- Dunbar SA (2006) Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin Chim Acta* 363: 71–82
- Francois P, Huyghe A, Charbonnier Y, Bento M, Herzig S, Topolski I et al. (2005) Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of *Staphylococcus aureus* isolates. *J Clin Microbiol* 43: 3346–3355
- Garaizar J, Rementeria A, and Porwollik S (2006) DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? *FEMS Immunol Med Microbiol* 47: 178–189
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, and Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303: 327–332
- Hamill M, Benn P, Carder C, Copas A, Ward H, Ison C, and French P (2007) The clinical manifestations of anorectal infection with lymphogranuloma venereum (LGV) versus non-LGV strains of *Chlamydia trachomatis*: a case-control study in homosexual men. *Int J STD AIDS* 18: 472–475
- Harmsen D, Claus H, Witte W, Rothganger J, Claus H, Turnwald D, and Vogel U (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol* 41: 5442–5448
- Huijsdens XW, van Dijke BJ, Spalburg E, van Santen-Verheuevel MG, Heck ME, Pluister GN et al. (2006) Community-acquired MRSA and pig-farming. *Ann Clin Microbiol Antimicrob* 5: 26
- Jonas D, Meyer HG, Matthes P, Hartung D, Jahn B, Daschner FD, and Jansen B (2000) Comparative evaluation of three different genotyping methods for investigation of nosocomial outbreaks of Legionnaires' disease in hospitals. *J Clin Microbiol* 38: 2284–2291
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35: 907–914
- Klint M, Fuxelius HH, Goldkuhl RR, Skarin H, Rutemark C, Andersson SG et al. (2007) High-resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence analysis. *J Clin Microbiol* 45: 1410–1414
- Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, and Musser JM (2001) Networks and tuberculosis: an undetected community outbreak involving public places. *Soc Sci Med* 52: 681–694
- Koek AG, Bovee LP, van den Hoek JA, Bos AJ, and Bruisten SM (2006) Additional value of typing Noroviruses in gastroenteritis outbreaks in Amsterdam, The Netherlands. *J Clin Virol* 35: 167–172

- Koeleman JG, Stoof J, Biesmans DJ, Savelkoul PH, and Vandenbroucke-Grauls CM (1998) Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. *J Clin Microbiol* 36: 2522–2529
- Kolader ME, Dukers NH, van der Bij AK, Dierdorp M, Fennema JS, Coutinho RA, and Bruisten SM (2006) Molecular epidemiology of *Neisseria gonorrhoeae* in Amsterdam, The Netherlands, shows distinct heterosexual and homosexual networks. *J Clin Microbiol* 44: 2689–2697
- Kremer K, Arnold C, Cataldi A, Gutierrez MC, Haas WH, Panaiotov S et al. (2005) Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. *J Clin Microbiol* 43: 5628–5638
- Lewis F, Hughes GJ, Rambaut A, Pozniak A, and Leigh Brown AJ (2008) Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 5: e50
- Lindstedt BA (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 26: 2567–2582
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95: 3140–3145
- Noller AC, McEllistrem MC, Shutt KA, and Harrison LH. (2006) Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. *J Clin Microbiol* 44: 374–377
- Pilcher CD, Wong JK, and Pillai SK (2008) Inferring HIV transmission dynamics from phylogenetic sequence relationships. *PLoS Med* 5: e69
- Reisig F, Kremer K, Amthor B, van Soolingen D, and Haas WH. (2005) Fast ligation-mediated PCR, a fast and reliable method for IS6110-based typing of *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 43: 5622–5627
- Schouls LM, van der Heide HG, Vauterin L, Vauterin P, and Mooi FR (2004) Multiple-locus variable-number tandem repeat analysis of Dutch *Bordetella pertussis* strains reveals rapid genetic changes with clonal expansion during the late 1990 s. *J Bacteriol* 186: 5496–5505
- Schouls LM, van der Ende A, Damen M., and van de Pol I (2006) Multiple-locus variable-number tandem repeat analysis of *Neisseria meningitidis* yields groupings similar to those obtained by multilocus sequence typing. *J Clin Microbiol* 44: 1509–1518
- Singh A, Goering RV, Simjee S, Foley SL, and Zervos MJ (2006) Application of molecular techniques to the study of hospital infection. *Clin Microbiol Rev* 19: 512–530
- Snell RG, and Wilkins RJ (1986) Separation of chromosomal DNA molecules from *C. albicans* by pulsed field gel electrophoresis. *Nucleic Acids Res* 14: 4401–4406
- Snipes KP, Hirsh DC, Kasten RW, Hansen LM, Hird DW, Carpenter TE, and McCapes RH (1989) Use of an rRNA probe and restriction endonuclease analysis to fingerprint *Pasteurella multocida* isolated from turkeys and wildlife. *J Clin Microbiol* 27: 1847–1853
- Spaargaren J, Schachter J, Moncada J, de Vries HJ, Fennema HS, Pena AS et al. (2005) Slow epidemic of lymphogranuloma venereum L2b strain. *Emerg Infect Dis* 11: 1787–1788
- Stene-Johansen K, Tjon G, Schreier E, Bremer V, Bruisten S, Nguui SL et al. (2007) Molecular epidemiological studies show that hepatitis A virus is endemic among active homosexual men in Europe. *J Med Virol* 79: 356–365
- Syvanen AC (2005) Toward genome-wide SNP genotyping. *Nat Genet* 37 Suppl: S5–S10
- Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH., and Swaminathan B (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 33: 2233–2239
- Tjon G, Xiridou M, Coutinho R, and Bruisten S (2007) Different transmission patterns of hepatitis A virus for two main risk groups as evidenced by molecular cluster analysis. *J Med Virol* 79: 488–494
- van Belkum A, Duim B, Regelink A, Moller L, Quint W, and van Alphen L. (1994) Genomic DNA fingerprinting of clinical *Haemophilus influenzae* isolates by polymerase chain reaction

- amplification: comparison with major outer-membrane protein and restriction fragment length polymorphism analysis. *J Med Microbiol* 41: 63–68
- van Belkum A, Kluytmans J, van Leeuwen W, Bax R, Quint W, Peters E et al. (1995) Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* 33: 1537–1547
- van Belkum A, Scherer S, van Alphen L, and Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* 62: 275–293
- van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK et al. (2007) Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* 13 Suppl 3: 1–46
- van den Braak N, Simons G, Gorkink R, Reijans M, Eadie K, Kremers K et al. (2004) A new high-throughput AFLP approach for identification of new genetic polymorphism in the genome of the clonal microorganism *Mycobacterium tuberculosis*. *J Microbiol Methods* 56: 49–62
- van der Zanden AG, Kremer K, Schouls LM, Caimi K, Cataldi A, Hulleman A et al. (2002) Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* complex isolates by introduction of new spacer oligonucleotides. *J Clin Microbiol* 40: 4628–4639
- van der Zee A, Verbakel H, van Zon JC, Frenay I, van Belkum A, Peeters M et al. (1999) Molecular genotyping of *Staphylococcus aureus* strains: comparison of repetitive element sequence-based PCR with various typing methods and isolation of a novel epidemicity marker. *J Clin Microbiol* 37: 342–349
- van Deutekom H, Supply P, de Haas PE, Willery E, Hoijng SP, Locht C et al. (2005) Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J Clin Microbiol* 43: 4473–4479
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31: 406–409
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Homes M et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23: 4407–4414
- Voss A, Loeffen F, Bakker J, Klaassen C, and Wulf M (2005) Methicillin-resistant *Staphylococcus aureus* in pig farming. *Emerg Infect Dis* 11: 1965–1966
- Ward H, Ison CA, Day SE, Martin I, Ghani AC, Garnett GP et al. (2000) A prospective social and molecular investigation of gonococcal transmission. *Lancet* 356: 1812–1817
- Williams JG, Kubelik AR, Livak KJ, Rafalski JA, and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18: 6531–6535
- Zhang Y, Lu L, Ba L, Liu L, Yang L, Jia M et al. (2006) Dominance of HIV-1 subtype CRF01_AE in sexually acquired cases leads to a new epidemic in Yunnan province of China. *PLoS Med* 3: e443