Brief description of frequently used typing methods

This document contains brief descriptions in a non-technical language of the principles of:

- PFGE
- MLVA
- Phage typing
- Sequence typing
- Typing of food- or waterborne viruses - Norovirus and hepatitis A virus
- Typing of food- or waterborne parasites – Giardia and Cryptosporidium

The document is intended to be of help to non-microbiologists working with FWD outbreak investigations. Please note that typing methods develop rapidly and that the information in this document may need future revisions. The current version was created April 2011.

What is PFGE?

PFGE is a typing method that is widely used for foodborne bacterial pathogens such as salmonella, campylobacter, listeria, VTEC and shigella. The principle is that the bacterial genome (DNA) is cut into typically 10-20 fragments that are separated by gel electrophoresis. Different clones of bacteria will (most likely) have their DNA cut differently. The method thus produces a banding pattern (a PFGE profile) based on the location of the cutting sites (restriction sites is the proper term) in the genome. A PFGE profile is a pattern of such DNA fragments separated according to size.

When performing PFGE, the circular bacterial DNA is treated with a particular restriction enzyme. These are protein structures that bind to particular sequences of normally 6 DNA letters (eg TCTAGA, but nowhere else in the DNA) and cut the DNA in two at these sites. The particular 6-letter sequences are only present a small number of times (eg 10 times) in the DNA and so the result is very large DNA fragments of varying lengths depending on which location of these sites the bacteria happened to have. To measure the size of the DNA fragments, the DNA is run out on a gel, separating the fragments by size using electric current (electrophoresis). The particular gel used is called a pulsed-field gel, because it uses a pulsating electrical field to obtain better separation of these very large DNA fragments, that otherwise tend to stick together.

The final result one gets is a pattern of bands (the DNA fragments) on a gel. This patterns is similar if two bacterial isolates are very similar or identical and otherwise different to varying degrees (if quite similar, the typists may talk of one or two "band-differences" and sometimes not be sure if the isolates are in fact very similar after all). To be sure that identical band patterns represent identical isolates, it's best to perform the analysis using different restriction enzymes (two, more rarely three enzymes may be used; frequently used enzymes include: *Xba*I, *Bln*I and *Sma*I). Also, it helps to know something about how often random isolates (for instance of a given salmonella serotype) will give the same pattern, because the variation in the distribution of the 6-letter DNA-sequences happen to be small for that serotype. So typing of unrelated (non-outbreak) isolates is good to have also for comparison. The method gives little information as too exactly how different isolates are, if they give different PFGE patterns. All you can say is that they are not the same (which is usually all you want to know in an outbreak situation).

To compare bandpatterns, a picture of the gel is taken and analysed using a special software (typically Bionumerics). Comparing results thus involves exchanging and comparing pictures. For PFGE, it's very important that conditions are the same when results from different laboratories are compared. That means same protocol (restriction enzyme, size marker, electrophoresis running condition), but also same equipment and reagents and training of technical staff.

This method is sometimes referred to as the gold standard of molecular typing and is the basis of the PulseNet system which has been very successful in the US. Among it's strengths is that it is widely applicable (many different types of bacteria) and that it distinguishes bacteria to a degree that is epidemiologically useful. Among the drawbacks are that it's a quite labour intensive method that usually takes three days from a plate culture to a profile.

In the papers included, there are several examples of what PFGE gels look like.

What is MLVA?

MLVA is a typing method that distinguishes different clones of related enteric bacteria by the presence of particular areas in the bacterial DNA.

The method is based on the presence of small groups of repeated DNA letters (e.g. CGTT repeated 10 times after each other) that sometimes exist in the bacterial DNA. These repeats have no apparent function for the bacteria, they just happen to be there and be useful for us. The number of times the groups of DNA letters are repeated will vary slightly even among fairly closely related strains of bacteria. The method then measures the number of repeats in a few well defined locations in the DNA, where such repeats are know to be present. The result of the analysis is simply these lengths. It could for instance be, if a MLVA method was measuring lengths at, say, five such DNA locations (called loci in DNA jargon): 5, 10, 18, 7 and 0.

To measure the number of repeats, the DNA areas under study are first amplified using PCR and primers containing specific dyes. This way one gets a series of DNA fragments where the DNA from the different locations have different colour coding. The fragments have sizes that reflect the number of repeats in the DNA. The exact sizes are measured by running them on a gel, separating them in size using electrical current. The protocols in current use involve using capillary electrophoresis for this. The different DNA fragments are distinguished within the same lane gel by their colour coding which are read using a laser.

MLVA has an important disadvantage: it is (at least up till now) not universal like PFGE but serotype dependent. That means that you for instance would need different types of MLVA methods for different salmonella serotypes. These MLVA methods have to be developed for each serotype and for some serotypes is has proven to be difficult to find a MLVA method that works (such as for *S*. Enteritidis). However, when it works, it generally works well. MLVA may distinguish related types of bacteria better than PFGE does. Also it's cheaper and produces results (a series of numbers) that can easily be communicated and compared between countries.

MLVA stands for multi locus VNTR analysis. VNTR stands for Variable Number of Tandem Repeat analysis. MLVA is also sometimes referred to as VNTR.

What is phage typing?

Phages are viruses that attack bacteria. When this happens the phage (bacterial virus) will copy itself within the bacteria and either kill the bacteria or stay in the bacteria in a dormant form. When this happens the phage will also be copied when the bacteria divides. Bacteria may thus have several phages sitting in them. Such phages will affect the ability of other phages to infect the bacteria. In other words, the ability to be infected by different phages varies between different strains of bacteria, even if they are fairly closely related. This fact forms the basis of phage typing.

The bacterial strains are grown and then subjected to attack by a series of different known phages. Some phages will kill the bacteria (this is clearly visible and therefore measurable) and others won't be able to kill a given bacteria. Depending on which groups of phages can kill or not kill a bacterial strain (the reaction pattern) the bacteria is given a number, the phage type.

Phage typing is most often used for the two common salmonella types, *S*. Enteritidis and *S*. Typhimurium. But phage typing systems also exist for some other salmonella serotypes and a few other bacteria. There are several such phage typing systems around. Probably the most widely used systems are those originally developed in England, by which *S*. Enteritidis isolates are called pt and then a number (e.g. *S*. Enteritidis pt4) and *S*. Typhimurium isolates often DT (Definitive Type) or U followed by a number (e.g. *S*. Typhimurium DT104).

Phage typing is not a DNA method (and is therefore sometimes said to be phenotypic rather than genotypic) and has been around for several decades. The phage type designations are therefore often well known and much used in the microbiological community. The method itself requires that the different phages are available and it's therefore a method that can generally only be performed at reference laboratories. Also it requires substantial technical expertise to perform.

What is sequence typing?

Sequence typing comprises several methods used for particular bacterial organisms. Here you take different strains and compare the exact sequences (the DNA letters) of a single gene. Not any gene can be used; the gene chosen will be one that is known to contain variation in its DNA letters, a variation that is known to reflect differences between strains that also make sense in epidemiological terms. For different organisms there are different preferred genes to study.

An example could be campylobacter, where *flaA*-SVR typing may be used. This is a sequence typing method that makes use of a particular short variable region (hence the name SVR) of about 150 DNA-letters in the *flaA* gene which encodes part of the bacterial flagella. It has been found that this region of the gene often varies between strains and that categorizing the isolates by *flaA*-SVR DNA sequence places them in epidemiologically relevant groups and that it can therefore be used for typing. Typically applying the method will mean amplifying this particular DNA region by PCR and analysing the DNA in various ways; one choice being to simply sequence the area (i.e. determine the exact order of the DNA letters). The results produced by such methods will say if two strains are (seemingly) identical and if not will also say something about to which degree the strains are different, i.e. how related they are.

Sometimes the principle is extended to comprise several genes. The method is then often called MLST, which stands for Multiple Locus (which here means gene) Sequence Typing. This may give very detailed information which may be valuable in scientific studies, but is generally not needed in an outbreak situation.

Typing of foodborne viruses

Norovirus and hepatitis A virus (HAV) are RNA viruses (their genome consist of RNA and not DNA). Norovirus and HAV may be detected in stool samples by PCR (the PCR involves two steps, the normal PCR and an initial special PCR that takes the RNA to DNA). The virus sequences may then be characterised and the viruses divided into different groups based hereon.

HAV is traditionally diagnosed by detection of specific antibodies in the blood following infection. In addition, HAV can often be detected by PCR in blood samples taken during the acute phase of the disease and these virus positive samples can be used for further typing assays.

Norovirus cannot be grown in culture, but is detected directly from stool samples, normally by PCR. In stools of acutely infected persons, there will often be very high amounts of virus present which makes detection feasible, whereas detection in foods or water very may be very difficult or impossible.

The viruses may be divided into different groups. Different PCR primers may be needed to detect viruses from different groups, so often a positive PCR detection will in itself help to classify the viruses into broad groups. Following detection by PCR, further characterization (typing) of these viruses can be done by determination of the sequence (DNA letters) of the whole or parts of the viral genome. However, the parts of the genome used for typing are most often not the same as the parts of the genome used for diagnostic testing. For diagnostic purposes genome regions with highly conserved nucleotide sequences are used, whereas the regions selected for typing usually have more variable sequences. This variability is the reason for the fact that typing tests (the PCR performed to obtain DNA for sequencing) often are less sensitive than the diagnostic tests and that a panel of different typing PCRs has to be applied before typing results can be obtained.

The sequenced obtained can be compared and used to say something about how related different viral isolates are and to place them into subgroups. Identical (or nearly identical) sequences from two different virus patient isolates will be a very strong indication that the patients were infected from the same source.

Typing of food- and waterborne parasites: Giardia & Cryptosporidium

Giardia and *Cryptosporidium* are both protozoan, eukaryotic parasites. In other words, they consist of a single cell which has the same basic lay-out as human cells; quite different (much larger and more complex) from bacterial cells.

Both Giardia and Cryptosporidium may be detected either by microscopy of stool sample material or directly by a PCR analysis on stool material.

Only one type of Giardia organism is recognised. The different names that are in use, *Giardia intestinalis*, *Giardia lamblia*, or *Giardia duodenalis*, are synonymous and denote the same organism. In contrast, cryptosporidiae, are divided into different species; currently 18 different species are recognised. Of these seven are considered as zoonotic with potential to infect humans. However, infections in humans is most often caused by one of only two species, *Cryptosporidium parvum* and *Cryptosporidium hominis*.

Both types of organisms may be subtyped using the techniques described above as 'sequence typing'. One or more particular genes are chosen, and parts hereof amplified by PCR and sequenced. Hereafter, the exact sequences (the DNA letters) of the gene(s) are compared. Identity means that two specimens are of the same subtype, variation can be used to classify the organisms into different subgroups. There is not currently a standardisation of which gene to use, so several classification systems exist. Also classification into different groups not based on sequencing, but on PCR-amplification of areas in the genome known to contain different numbers of repeat areas, may be used. These areas are then distinguished based on their varying size.

In contrast to bacteria, these organisms can not easily be grown in culture. Therefore, subtyping of *Giardia* and *Cryptosporidium* may be technically challenging, simply because of low amount of parasitic material that can make it difficult to run the PCRs.