

REVIEW ARTICLE

Bacterial strain typing in the genomic era

Wenjun Li, Didier Raoult & Pierre-Edouard Fournier

Unité des Rickettsies et pathogènes émergents, CNRS UMR 6236, IRD 198, Faculté de médecine, Université de la Méditerranée, Marseille, France

Correspondence: Pierre-Edouard Fournier, Unité des Rickettsies et pathogènes émergents, CNRS UMR 6236, IRD 198, Faculté de médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France. Tel.: +33 04 91 38 55 17; fax: +33 04 91 83 03 90; e-mail: pierre-edouard.fournier@medecine.univ-mrs.fr

Received 23 June 2008; revised 26 March 2009; accepted 1 April 2009.
Final version published online 6 May 2009.

DOI:10.1111/j.1574-6976.2009.00182.x

Editor: Ferran Garcia-Pichel

Keywords

genotyping; bacteria; genome; DNA.

Abstract

Bacterial strain typing, or identifying bacteria at the strain level, is particularly important for diagnosis, treatment, and epidemiological surveillance of bacterial infections. This is especially the case for bacteria exhibiting high levels of antibiotic resistance or virulence, and those involved in nosocomial or pandemic infections. Strain typing also has applications in studying bacterial population dynamics. Over the last two decades, molecular methods have progressively replaced phenotypic assays to type bacterial strains. In this article, we review the current bacterial genotyping methods and classify them into three main categories: (1) DNA banding pattern-based methods, which classify bacteria according to the size of fragments generated by amplification and/or enzymatic digestion of genomic DNA, (2) DNA sequencing-based methods, which study the polymorphism of DNA sequences, and (3) DNA hybridization-based methods using nucleotidic probes. We described and compared the applications of genotyping methods to the study of bacterial strain diversity. We also discussed the selection of appropriate genotyping methods and the challenges of bacterial strain typing, described the current trends of genotyping methods, and investigated the progresses allowed by the availability of genomic sequences.

Introduction

Genetic diversity can ultimately explain most phenotypic variability in bacteria, such as geographic distribution, host specificity, pathogenicity, antibiotic resistance, and virulence. As bacterial strains pose ever greater challenges to human health, including increased virulence and transmissibility, resistance to multiple antibiotics, expanding host spectra, and the possibility of genetic manipulation for bioterrorism, identifying bacteria at the strain level is increasingly important in modern microbiology (Fournier *et al.*, 2004).

The intraspecies diversity of bacteria mainly results from three genetic events: horizontal gene transfer, gene loss or acquisition, and recombination. The frequency of these three events makes the investigation of intraspecies diversity quite complicated (Fraser-Liggett, 2005). Although whole-genome sequencing, followed by genome comparison, is in theory an ideal way to elucidate the genetic variability within a bacterial species, it remains expensive and labor intensive (Field *et al.*, 2004). Bacterial strain typing, characterizing a number of strains in detail and ascertaining whether they are derived from a single parental organism, is a way to identify bacteria at the strain level and to uncover the genetic diversity underlying important phenotypic characteristics.

Theoretically, there are two distinct typing systems: phenotyping and genotyping. Bacterial phenotypes, determined by the morphology of colonies on various culture media, biochemical tests, serology, killer toxin susceptibility, pathogenicity, and antibiotic susceptibility, are not variable enough for discriminating between closely related strains. In the genomic era, the scientific basis for the identification and subtyping of microorganisms has shifted to genetic methods. More recently, genotyping, which refers to the discrimination of bacterial strains based on their genetic content, has become widely used for bacterial strain typing due to its high resolution. The genetic profile of a given strain generated by a specific genotyping method can be as unique as a fingerprint. Thus, genotyping is also referred to as DNA fingerprinting.

Current bacterial strain typing methods may be classified into three main categories: DNA banding pattern-, DNA sequencing-, and DNA hybridization-based methods. DNA banding pattern-based genotyping methods discriminate the studied strains based on differences in the size of the DNA bands (fragments) generated by amplification of genomic DNA or by cleavage of DNA using restriction enzymes (REs). DNA sequencing-based genotyping methods generate the original sequence of nucleotides and discriminate among

bacterial strains directly from polymorphisms in their DNA. DNA hybridization-based methods are mainly referred to as DNA macroarray and microarray studies. In this technique, bacterial strains are discriminated by analyzing the hybridization of their DNA to probes of known sequences. With the exception of genome sequencing, the discriminatory ability of genotyping methods is species dependent.

Current methods for bacterial strain typing

DNA banding pattern-based methods (Fig. 1)

DNA bands (fragments) can be generated by digestion of DNA with REs, DNA amplification, or by a combination of both. DNA amplification generates billions of copies of a genomic fragment and has many advantages, including sensitivity, speed, and applicability to a wide array of human specimens and environmental samples (Lo & Chan, 2006). REs precisely recognize and cut target DNA at a defined sequence, which makes enzymatic restriction an effective tool (Blakesley, 1987).

Enzymatic restriction

Pulsed-field gel electrophoresis (PFGE)

PFGE is an electrophoretic technique used to separate large DNA molecules (10 kb–10 Mb). In a conventional constant electric field, DNA molecules > 20 kb show the same mobility, making it impossible to differentiate between these

DNA molecules. By applying alternating electric fields at different angles, however, PFGE can separate large DNA molecules in a flat agarose gel (Schwartz & Cantor, 1984; Herschleb *et al.*, 2007) (Fig. 1). REs with uncommon recognition motifs are used in PFGE to generate large DNA fragments (Table 1) and the banding patterns of PFGE in a group of strains reflect DNA polymorphism at the RE recognition sites. PFGE provides a high-resolution, macro-restriction analysis at the genome level, leading it to be considered as the ‘gold standard’ for subtyping many bacteria (Tenover *et al.*, 1995; Gerner-Smidt *et al.*, 2006).

The choice of RE is one of the most important factors in determining the PFGE banding pattern because the cleavage site of each RE is unique. REs with long, infrequently occurring recognition motifs may provide higher resolution in PFGE (Chen *et al.*, 2005). *In silico* searches of completed bacterial genome sequences to detect recognition sites of rare-cutting REs proved to be an effective way to optimize the discriminatory power and reduce the cost of PFGE in the case of typing *Bordetella pertussis* (Lee *et al.*, 2006). A website dedicated to *in silico* RE analysis of complete bacterial genomes was developed by Bikandi and colleagues. Their strategy allows the calculation of the theoretical number and size of restriction fragments generated by different REs and provides a graphically detailed restriction profile when the number of fragments is 50 or fewer (Bikandi *et al.*, 2004) (Table 1). Nevertheless, PFGE results should be validated by comparison with epidemiological traits and other genotyping results. For example, genotyping *Bartonella henselae*

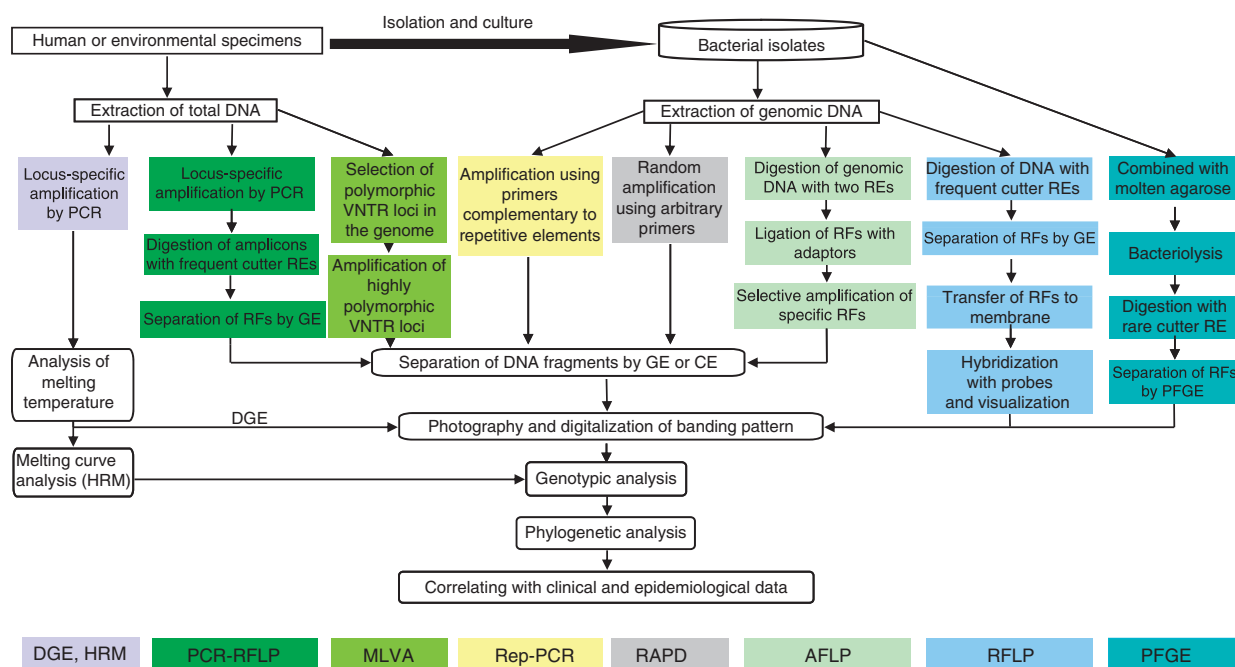


Fig. 1. Flow chart of DNA banding pattern-based genotyping methods for bacterial strains.

Table 1. Comparison of DNA banding pattern-based genotyping methods used in bacterial strain typing

Methods	Bacterial culture required		Primers needed	DNA fragments		Main reagents	Main equipments*
	REs			Size	Resolution		
PFGE	Yes	Rare cutters	No	Large (> 20 kb)	PFGE	REs and proteases	Gel box, switching unit, cooler, power supply
RFLP	Yes	Frequent cutters	No	Moderate (< 10 kb)	GE followed by Southern blotting	REs and probes	Southern transfer
Ribotyping	Yes	Frequent cutters	No	Moderate (< 10 kb)	GE followed by Southern blotting	REs and probes	Southern transfer
PCR-RFLP	No	Frequent cutters	Flanking target fragment	Small (< 5 kb)	GE	<i>Taq</i> , primers and REs	Thermal cycler
AFLP	Yes	A pair of REs	Selective primers [†]	Very small (< 1 kb)	GE or CE	REs, adaptors, T4, ligase, <i>Taq</i> , primers	Thermal cycler
REP-PCR	Yes	No	Complementary to repetitive elements	Small (< 5 kb)	GE or CE	<i>Taq</i> and primers	Thermal cycler
RAPD	Yes	No	Arbitrary primers	Small (< 5 kb)	GE or CE	<i>Taq</i> and primers	Thermal cycler
MLVA	No	No	Flanking tandem repeats	Very small (< 1 kb)	GE or CE	<i>Taq</i> and primers	Thermal cycler
DGE	No	No	Flanking target fragment	Small (< 5 kb)	Denaturing gel	<i>Taq</i> and primers	Thermal cycler, specific electrophoresis equipment
HRM	No	No	Flanking target fragment	Small (< 5 kb)	Melting curve analysis	<i>Taq</i> and primers	Real-time thermal cycler

*Conventional gel electrophoresis, photography, and digitalizing equipment are not included.

[†]Adapter sequences at 5' end, extending with a variable number of 3' nucleotides (usually one to three).

GE, gel electrophoresis; CE, capillary electrophoresis.

using NotI-based PFGE proved to be much more discriminatory than using four other REs: SmaI, ApaI, Eco52I, and XmaJI. However, NotI-based PFGE types did neither match other PFGE types nor genotypes obtained using multilocus sequence typing, indicating that NotI is not an appropriate RE for PFGE typing of *B. henselae* (Arvand & Viezens, 2007). Therefore, the choice of a suitable RE is crucial to generate reliable PFGE types.

PFGE is now widely used in epidemiology, microbiology, and evolutionary biology. The rapid accumulation of PFGE patterns necessitates the standardization of protocols and criteria for bacterial typing, as well as the development of online databases for global comparison of PFGE patterns (Ransom & Kaplan, 1998; Herschleb *et al.*, 2007). The largest PFGE database, PulseNet, currently tracks four foodborne bacteria: *Escherichia coli* O157:H7, nontyphoidal *Salmonella*, *Shigella*, and *Listeria monocytogenes* (Ransom & Kaplan, 1998; Ribot *et al.*, 2006; Swaminathan *et al.*, 2006) (Table 1). As proposed by Tenover and colleagues, bacterial isolates yielding the same PFGE profile are considered as belonging to the same strain. Isolates that differ by a single genetic event, which is reflected as a difference of one to three bands, are considered as being 'closely related,' and isolates differing by four to six bands, likely representing two independent genetic events, are considered as being 'possibly related.' Bacterial isolates containing six or more band differences

(representative of three or more genetic changes) are considered as being 'unrelated' (Tenover *et al.*, 1995).

However, although widely used, PFGE has several limitations. This method is time and labor consuming and lacks reproducibility and interlaboratory comparability. In addition, it requires high-quality DNA, is poorly applicable to human or environmental samples, and may lack resolution power to distinguish bands of nearly identical size (Davis *et al.*, 2003). Other drawbacks include the risk of laboratory-acquired infection due to prolonged handling of bacterial strains before treating with proteases and REs, and many other factors such as concentration of DNA in the agarose plugs, amount of agarose in the gel, electrophoresis voltage, gel temperature, and buffer strength, which may also influence patterns (Chung *et al.*, 2000).

Restriction fragment length polymorphism (RFLP)

RFLP analysis is one of the first techniques to be widely used for detection of variations in DNA sequence (Thibodeau, 1987; Todd *et al.*, 2001). As its name implies, RFLP measures the size of restriction fragments separated by conventional agarose gel electrophoresis. The general procedure for RFLP analysis is shown in Fig. 1. Digestion of genomic DNA with frequently cutting REs may produce hundreds of short restriction fragments, making clear separation of fragments

difficult using agarose gel electrophoresis (Table 1). However, RFLP analysis can be greatly simplified by subjecting the partial restriction fragments to Southern blotting with labeled probes (Southern, 1975).

In bacteriology, if two strains differ in the distance between cleavage sites for a particular RE, the length of the restriction fragments is different between the strains. The similarity of the generated patterns of restriction fragments can be used to differentiate strains and to analyze the genetic relatedness (Busse *et al.*, 1996). RFLP patterns are mainly determined through the specific combination of REs and nucleic acid probes. RFLP analysis using probes derived from the insertion element, *IS6110*, has been considered as the 'gold standard' method for typing *Mycobacterium tuberculosis* complex (MTC) bacteria (Otal *et al.*, 1991; Mostrom *et al.*, 2002; Kanduma *et al.*, 2003). These bacteria were also subtyped using RFLP analysis based on polymorphic GC-rich repeat sequences (Poulet & Cole, 1995; Mostrom *et al.*, 2002; Kanduma *et al.*, 2003).

Ribotyping, a variation of RFLP analysis, uses rRNA probing. Bacterial rRNA operons comprise a family of highly conserved genes, each of which is flanked by more variable DNA regions. Sequence variations in the flanking restriction sites result in different RFLP banding patterns when probed with conserved domains of the 16S or 23S rRNA genes (Bingen *et al.*, 1994; Harvey & Minter, 2005). The banding patterns of DNA fragments corresponding to the relevant rRNA are named ribotypes. An advantage of ribotyping is that it enables analysis without prior knowledge of genomic DNA sequence, because rRNA operons are universal. In addition, the results of ribotyping are easier to interpret because fewer fragments are produced (Bingen *et al.*, 1994; Harvey & Minter, 2005). An automated apparatus for ribotyping is commercially available, with the RiboPrinterTM Microbial Characterization System (Qualicon, Wilmington, DE). This system provides highly reproducible and standardized ribotyping data. Ribobank is a ribotyping database comprising numerous bacterial ribotypes generated by Qualicon's RiboPrinter approach (Table 2).

Because it does not require expensive equipment, RFLP analysis is cost effective except for automatic ribotyping that requires a relatively expensive kit. Among limitations, RFLP analysis requires large amounts of high-quality genomic DNA, which can limit its application in many cases. In addition, it is time and labor consuming, especially when coupled with Southern blotting, and involves detection systems that use either radioisotopes or complex biochemistry.

DNA amplification

Arbitrarily primed PCR (AP-PCR)

AP-PCR is a variation of classic PCR for random amplification of unknown genomic regions using arbitrary primers.

Therefore, it is also termed random amplification of polymorphic DNA (RAPD) (Welsh & McClelland, 1990; Williams *et al.*, 1990). Using short single primers (usually < 10 bp) combined with low annealing temperatures, AP-PCR enables genomic DNA to be amplified at multiple loci, which generates different-sized amplicons (Table 1, Fig. 1). Unlike traditional PCR analysis, AP-PCR does not require prior knowledge of the target DNA: the identical 10-mer primers may or may not amplify a segment of DNA, depending on the positions that are complementary to the primer sequences (Power, 1996) (Fig. 1). For example, no fragment is produced if primers anneal too far apart or if the 3' ends of the primers fail to face each other. If a mutation has occurred in the template DNA at a site that was previously complementary to the primer, a PCR product may not be produced, and a different pattern of amplified DNA fragments is observed on the gel.

Based on the observation that the number of fragments generated by AP-PCR correlates with the number of arbitrary primer annealing sites in the target genome sequence, Li *et al.* (2006a) proposed a statistical method and developed a computer program named ERAPD to determine the number of annealing sites and amplicons for RAPD typing (Table 2). Some arbitrary primers, namely the universal rice primers extracted from a repetitive sequence in the rice genome (Kang *et al.*, 2002; Jin *et al.*, 2006) and the highly variable penta-GTG oligonucleotide [(GTG)₅] (Matsheka *et al.*, 2006), have frequently been used for RAPD analysis of *Salmonella enterica* and *Campylobacter concisus*, respectively. 'RAPD-GENERATOR' a Java-based program is also useful for generating arbitrary primers (Table 2).

As a PCR-based genotyping method, AP-PCR is inexpensive, fast, and sensitive. It has widely been used to study the genetic variability of many bacterial species, including important human pathogens (Power, 1996; Jin *et al.*, 2006). However, the reproducibility of this method remains a challenge, which hinders comparison of AP-PCR patterns between and within laboratories (Power, 1996; Tyler *et al.*, 1997). In addition, several factors may influence AP-PCR results, such as sequence and annealing temperature of arbitrary primers, DNA template purity and concentration, and PCR equipment and reagents (Tyler *et al.*, 1997). Optimization and standardization of the procedure may create sufficiently reliable conditions for using AP-PCR in epidemiological typing of an expanding range of bacteria (Mortimer & Arnold, 2001; Mostrom *et al.*, 2002). The Ready-To-Go RAPD analysis beads (Amersham Pharmacia Biotech) provide a strictly consistent and uniform source of *Taq* polymerase, dNTPs, and buffer conditions along with a consensus protocol for RAPD analysis. This approach may prove useful for standardizing and optimizing this genotyping method (Hyytia *et al.*, 1999; Vogel *et al.*, 1999; Chansiripornchai *et al.*, 2000).

Table 2. Useful web resources for bacterial genotyping

Name	Website	Description
<i>Tools</i>		
REBASE enzymes	http://rebase.neb.com/rebase/rebase.html	Recognizing and cutting motifs of RES
ALFIE	http://www.hpa-bioinfotools.org.uk/cgi-bin/ALFIE/index.cgi	AFLP fragments prediction
AFLP in SILICO	http://bioinformatics.psb.ugent.be/webtools/aflpinsilico/	Prediction of AFLP fingerprinting profile
RAPD-generator	http://www2.uni-jena.de/biologie/mikrobio/tipps/rapid.html	Arbitrary primers generator for RAPD analysis
eRAPD	ask Dr Li via e-mail: lishihengt@nwsuaf.edu.cn	Determine the number of annealing sites and amplicons of RAPD by given primer
AFLP Manager	http://server.ispa.cnr.it/AFLPManager/	Standardization of AFLP and their integration with other biological data
<i>In silico</i> genotyping	http://insilico.ehu.es/	Prediction of fingerprinting profile of PFGE, AFLP, RFLP, PCR-RFLP
<i>Databases</i>		
GenBank	http://www.ncbi.nlm.nih.gov	The largest DNA sequence database including genome and gene sequences
GOLD	http://www.genomesonline.org/	Genome online database including ongoing genome sequencing projects
PFGE	http://www.cdc.gov/pulsenet/	PFGE of foodborne bacteria (CDC)
	http://www.pulsenet-europe.org/	PFGE of foodborne bacteria (Europe)
	http://www.mri.sari.ac.uk/bacteriology-current-project-01-02.asp	PFGE of mycobacteria
	http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1195733756584?p=1160994272616	PFGE of <i>Pseudomonas aeruginosa</i>
	http://www.fsis-pfge.org/index.html	PFGE of <i>Escherichia coli</i>
RFLP	http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1195733788576?p=1160994272616	PFGE of <i>Acinetobacter baumannii</i>
	http://www.hpa.org.uk/cfi/bioinformatics/salm_gene/salm_genedb.htm	PFGE of <i>Salmonella</i> sp.
	http://www.shigatox.net/cgi-bin/mlst7/rflpdb	RFLP of pathogenic <i>Escherichia coli</i>
Ribotyping	http://www.ewi.med.uu.nl/gene/ribobank.htm , and Ribotyping of <i>Clostridium difficile</i>	Ribotyping
AFLP	http://www.hpa-bioinformatics.org.uk/cgi-bin/ALFIE/index.cgi	AFLP of <i>Legionella pneumophila</i> AFLP of <i>Mycobacterium tuberculosis</i>
Spoligotyping	http://www.mbovis.org/index.php	Spoligotyping of <i>Mycobacterium tuberculosis</i>
MLVA	http://www.pasteur-guadeloupe.fr/tb/bd_myco.html	Spoligotyping of <i>Mycobacterium tuberculosis</i>
	http://www.umcutrecht.nl/subsite/MLVA/	MLVA of <i>Enterococcus faecium</i>
	http://minisatellites.u-psud.fr/	Tandem repeat database of bacteria and MLVA tools
	http://cagt.bu.edu/page/TRDB_about	Tandem repeat database
MLST	http://insilico.ehu.es/	Microsatellite repeats database
	http://vntr.csie.ntu.edu.tw/	Bacteria tandem repeat database
	http://www.mlst.net	MLST of 23 bacterial species
MST	http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php	MLST of <i>Legionella pneumophila</i>
	http://www.shigatox.net/cgi-bin/mlst7/dbquery	MLST of pathogenic <i>Escherichia coli</i>
	http://ifr48.timone.univ-mrs.fr/portail2/	MST of four bacterial species

Repetitive sequencing-based PCR (REP-PCR)

A series of naturally occurring repetitive DNA sequences are dispersed in multiple copies throughout bacterial genomes (Gilson *et al.*, 1984; Stern *et al.*, 1984; Hulton *et al.*, 1991; Lupski & Weinstock, 1992; Koeuth *et al.*, 1995). Although the functions of these interspersed repetitive DNA elements remain unknown, their presence is useful for DNA fingerprinting of bacteria. In REP-PCR, primers complementary to those interspersed repetitive consensus sequences are used to

amplify DNA fragments between repetitive elements (Versalovic *et al.*, 1991; de Bruijn, 1992) (Fig. 1). Three families of repetitive sequences: the 35–40-bp repetitive extragenic palindromic (REP) sequence (Gilson *et al.*, 1984; Stern *et al.*, 1984), the 124–127-bp enterobacterial repetitive intergenic consensus (ERIC) sequence (Hulton *et al.*, 1991; Lupski & Weinstock, 1992), and the 154-bp BOX element sequences (Koeuth *et al.*, 1995), have frequently been used in REP-PCR assays. The corresponding protocols are referred to as REP-PCR, ERIC-PCR, and BOX-PCR genomic fingerprinting, respectively.

Detection of DNA fragments is obtained by agarose gel electrophoresis or capillary electrophoresis. The optimization of electrophoresis conditions (electrokinetic injection and applied voltage) can dramatically increase the resolution of amplified DNA fragments up to 1000 bp (Sciacchitano, 1998).

Because of the low cost of the materials and the rapidity, ease of use, and low labor intensity, REP-PCR may be a valuable method for bacterial strain typing. In contrast to AP-PCR, REP-PCR is more reproducible because specific primers are used for amplification. REP-PCR has been shown to be beneficial for typing eukaryotic and prokaryotic organisms, including bacteria of epidemiological significance, i.e. *Acinetobacter* spp. (Snelling *et al.*, 1996), *Streptococcus pneumoniae* (Versalovic *et al.*, 1993), *M. tuberculosis* (Cangelosi *et al.*, 2004), *Clostridium difficile* (Northey *et al.*, 2005), and *L. monocytogenes* (Chou & Wang, 2006). A commercial kit, the REP-PCR DiversiLab microbial typing system (Spectral Genomics Inc., Houston, TX), is reported to be more convenient than manual methods (Healy *et al.*, 2005; Shutt *et al.*, 2005). The online DIVERSILAB software not only provides standardized comparisons among isolates almost instantaneously, but also generates user-friendly customized reports and provides a user-specific data storage and retrieval system (Healy *et al.*, 2005). In addition, REP-PCR genomic fingerprinting protocols can be performed on whole cells of some species, obviating the need for DNA extraction (Woods *et al.*, 1993; Snelling *et al.*, 1996). However, REP-PCR has disadvantages similar to other PCR-based assays, including the potential for contamination, artifacts, and the need for multiple controls.

Multiple-locus variable number tandem repeat analysis (MLVA)

Variable number tandem repeats (VNTR) are 'head-to-tail' (tandemly) repeated DNA sequences that vary in copy number and are dispersed widely in human and bacterial genomes (Lupski & Weinstock, 1992; van *et al.*, 1998; Vergnaud & Denoeud, 2000). In bacterial genomes, VNTR loci were found in noncoding regions as well as in genes. These VNTR can be an important source of genetic polymorphism for strain typing because of their rapid evolution (van *et al.*, 1998, 2001; Ogata *et al.*, 2000; Fournier *et al.*, 2004). The number of tandem repeats per locus may vary dramatically between strains within a given species. MLVA is a PCR-based genotyping method based on the polymorphic analysis of multiple VNTR loci on the chromosome (Lindstedt, 2005; van, 2007). For each VNTR locus, the number of repeats can be determined by PCR amplification using primers complementary to the well-conserved sequences flanking the tandem repeats. The fragment size varies with regard to the size and number of repeat units, and the banding patterns can be analyzed using a software to reveal

the genotype and infer the phylogenetic relationships (Keim *et al.*, 2000; Tenover *et al.*, 2007) (Fig. 1). By fluorescently labeling PCR amplicons with different dye colors and separating them by capillary electrophoresis in an automated sequencer, MLVA can be performed in a single PCR tube (multiplex PCR), which dramatically improves its efficiency (Lindstedt, 2005; Lista *et al.*, 2006).

Scanning the bacterial genome and extracting VNTR loci are the first steps in MLVA. Tandem repeat-finding programs, such as TANDEM REPEAT FINDER (Benson, 1999), the Tandem Repeat database, and the Microsatellite Repeats database provide a series of bioinformatic tools for searching bacterial genomes for potential VNTR loci (Table 2). Databases and web-based tools for VNTR searching, primer picking, and phylogenetic analysis facilitate the implementation of MLVA (Table 2) (Denoeud & Vergnaud, 2004; Grissa *et al.*, 2007).

Since its first use to type bacteria in 2000, MLVA has proven to be a high-resolution method for discrimination of many bacteria. It is now regarded as a reference typing method for many bacterial species, such as *Francisella tularensis* (Farlow *et al.*, 2001; Johansson *et al.*, 2004), *Bacillus anthracis* (Keim *et al.*, 2000; Hoffmaster *et al.*, 2002; Lista *et al.*, 2006), *Yersinia pestis* (Klevytska *et al.*, 2001), and *M. tuberculosis* (Le *et al.*, 2002). MLVA has also been applied to other important human pathogens such as methicillin-resistant *Staphylococcus aureus* strains (Tenover *et al.*, 2007), *Burkholderia pseudomallei* (U'Ren *et al.*, 2007), and *C. difficile* (van den Berg *et al.*, 2007). Because of the rapid evolution of VNTR loci in bacterial genomes, MLVA has the same or more discriminatory power than most other typing methods (Johansson *et al.*, 2004; Lindstedt, 2005; Liao *et al.*, 2006; Tenover *et al.*, 2007).

The stability of VNTRs is mainly associated with the length of repetitive unit, copy number of repetitive unit, and the purity (Legendre *et al.*, 2007). A program, SERV, is now available for the selection and comparison of suitable VNTR for genotyping based on scoring the length of repetitive unit, copy number of repetitive unit, and the purity (Legendre *et al.*, 2007) (<http://hulsweb1.cgr.harvard.edu/SERV/>).

Although MLVA is a rapid, easy to perform, inexpensive, and reproducible genotyping method with high resolution, VNTR loci may evolve too quickly to provide reliable phylogenetic relationships among closely related strains. In *Mycobacterium leprae*, variation in the VNTR pattern was observed not only between isolates of *M. leprae* but also between biopsies from the same patient (Monot *et al.*, 2008). Therefore, MLVA is unsuitable for long-term epidemiological surveillance, though it may be useful for tracking outbreaks of bacterial infections (Lindstedt, 2005). In addition, MLVA showed less discriminatory power than multilocus sequence typing (MLST) and PFGE for *Enterococcus faecium* (Werner *et al.*, 2007). In addition, VNTR loci are not always common in bacterial genomes, such as in *Mycoplasma*

hyopneumoniae (Minion *et al.*, 2004), which limits the implementation of MLVA. The size difference in a VNTR locus may not always reflect the real number of tandem repeats because insertions or deletions in the amplified region can also give rise to the same size difference. Therefore, sequencing of the amplicons is necessary in this case. However, with the increase in the number of bacterial genome sequences available for a particular species, it may be possible to improve MLVA through the rational selection of suitable tandem repeat loci and primer design (Johansson *et al.*, 2004; Lindstedt, 2005).

Denaturing gel electrophoresis (DGE)

In DGE, PCR amplicons with the same length are separated by electrophoresis in polyacrylamide gels in a sequence-dependent manner. The increasing gradient of denaturing components along the polyacrylamide gel opens double-stranded amplicons into single-stranded DNA through melting domains. The melting temperature of these domains is sequence-specific and decreases their mobility. Thus, different sequences will result in different origins of melting domains and consequently in different positions in the gel (Muyzer *et al.*, 1993). Such a gradient is obtained using either denaturing chemicals for denaturing gradient gel electrophoresis (DGGE), or heat for temperature gradient gel electrophoresis (TGGE) and temporal temperature gradient electrophoresis (TTGE) (Fromin *et al.*, 2002). 16S rRNA gene is the most frequent target gene for DGE analysis because it exists in all bacteria and can easily be amplified without prior knowledge of studied strains. Applying statistical method makes DGE fingerprinting technique a promising tool (Fromin *et al.*, 2002). Numerous samples can be analyzed simultaneously, allowing the monitoring of microbial communities.

Constant denaturant capillary electrophoresis (CDCE) is similar to DGGE but has better resolution and can be quantitative (Lim *et al.*, 2001). This method consists of competitive quantitative PCR with group-specific primers targeting a variable region of 16S rRNA gene, followed by separation and quantification of the amplicons in the capillary containing high-molecular-weight linear polyacrylamide gel. CDCE can resolve sequences that differ by as little as a single base pair, and quantification of sequences is extremely sensitive due to the use of a laser-based detection system (Lim *et al.*, 2001; Thompson *et al.*, 2004).

DGE has been widely used in exploring genetic diversity of bacteria since its first application in 1993 (Muyzer *et al.*, 1993). An advantage of DGGE is that selected bands can be sequenced. Thus, the presence of a particular phylotype can be monitored in the samples studied. In addition, DGE can be applied directly to environmental samples or specimens because it is PCR based. However, DGGE requires a

specific electrophoresis equipment and it is not possible to run several gels per run.

High-resolution melting (HRM) analysis

Like DGE, HRM analysis is aimed at discriminating DNA alleles through melting temperature analysis. HRM uses real-time PCR amplification and melting curve analysis. The melting profile of a PCR product depending on its GC content, length, sequence, and heterozygosity, is conveniently monitored with saturating dyes that fluoresce in the presence of double-stranded DNA. HRM is not a banding pattern-based method *sensu stricto* but we classified it in this category of methods because it is based on PCR amplification and allows detection of sequence variants without sequencing or hybridization procedures (Gundry *et al.*, 2003). Using HRM, single nucleotide polymorphisms (SNPs) can be genotyped without probes and more complex regions can be typed with unlabeled hybridization probes. HRM can be combined with various PCR strategies for bacterial strain genotyping. For *Mycoplasma synoviae*, HRM was used to detect 10 sequence variants in the *vlhA* gene (Jeffery *et al.*, 2007). Using the clustered, regularly interspaced short palindromic repeats (CRISPR) locus of *Campylobacter jejuni* as target, HRM exhibited a discriminatory power comparable to that of the gold standard genotyping method, i.e. PFGE (Price *et al.*, 2007). When combined to amplification of VNTRs, HRM provided similar results to MLVA for *B. anthracis* (Fortini *et al.*, 2007). The advantages of HRM are its homogeneity, rapidity (1–5 min), applicability to environmental samples, and the use of closed tubes. It has a similar or superior sensitivity and specificity to methods that require physical separation.

However, HRM requires real-time PCR equipment and may not be equally discriminatory for all bacterial species. In addition, comparability of results among laboratories may be limited.

Enzymatic digestion following DNA amplification

PCR-RFLP

PCR-RFLP, which involves RFLP analysis of a specific locus amplified by PCR, overcomes the shortcomings of RFLP (Wichelhaus *et al.*, 2001) (Fig. 1). Because this method is a PCR-based genotyping method, it can be used with DNA taken directly from human specimens and environmental samples. In addition, the limited number of restriction fragments resulting from RE digestion of the PCR amplicons can be separated and visualized directly by gel electrophoresis without the need for probe hybridization. PCR-RFLP analysis with capillary electrophoresis (PRACE) shows dramatically higher resolution than conventional PCR-RFLP (Ho *et al.*, 2004; Chang *et al.*, 2007b). Therefore, PCR-RFLP

is a simple, rapid, and nonradioactive approach to detect DNA polymorphism. It has been used frequently for typing a variety of bacteria including *Brucella* species (Al *et al.*, 2005) and *M. tuberculosis* (Hayward, 1995; Cohn & O'Brien, 1998).

The fact that genetic information in PCR-RFLP comes from a single locus often limits the discriminatory power of this technique. Multilocus-based PCR-RFLP analysis may enhance the discriminatory power.

Amplified fragment length polymorphism (AFLP)

AFLP, which was first described by Vos and Zabeau in 1993, combines the accuracy of RE analysis with the precision of PCR, making it a highly sensitive method for detecting DNA polymorphism (Vos *et al.*, 1995). In AFLP analysis, digestion of genomic DNA with two REs (usually *MseI* and *EcoRI*) and ligation of restriction fragments with end-specific adapters, is followed by selective amplification. The regions are amplified with primers consisting of the adapter sequences at the 5' end and extended with a variable number of 3' nucleotides (usually 1–3 bp) chosen by the user (Fig. 1). DNA fragments can be detected by conventional gel electrophoresis or automated DNA sequencer, as in other DNA banding pattern-based genotyping methods (Vos *et al.*, 1995; Mueller & Wolfenbarger, 1999; Zhao *et al.*, 2000) (Fig. 1). The fluorescent AFLP (FAFLP) method uses fluorescently labeled primers for PCR amplification and an automated DNA sequencer for fragment detection. FAFLP shows a much higher resolution than conventional AFLP detection systems, with the ability to detect size differences as small as 1 bp (Lindstedt *et al.*, 2000a; Zhao *et al.*, 2000; Mortimer & Arnold, 2001). Fragments are usually scored as either present or absent, and the presence or absence of bands in the second sample is referred to as polymorphism. Computer analysis of the number and size of fragments generated from each strain allows phylogenetic analysis (Lindstedt *et al.*, 2000b; Mortimer & Arnold, 2001).

The discriminatory power of AFLP analysis is mainly determined by the combination of REs and the number of selective nucleotides in the primers. *In silico* analysis of bacterial genomic sequences can be used to select the most informative combinations of ERs and primer sequences (Table 2) (Keto-Timonen *et al.*, 2003; Rombauts *et al.*, 2003; Bikandi *et al.*, 2004; Kivioja *et al.*, 2005).

AFLP offers great flexibility in the number of loci that can be amplified simultaneously in one PCR reaction. Typically, 50–100 restriction fragments are coamplified in one fingerprint, depending on the complexity of the genome, the primers used, and the resolution of gel electrophoresis. Thus, relatively few primer pairs are used to visualize a large number of loci in AFLP analysis. In comparison with other genotyping methods, AFLP has been described as being as discriminatory as PFGE, RAPD, and RFLP, and more discriminatory than

MLST (Zhao *et al.*, 2000; Feberwee *et al.*, 2005; Gzyl *et al.*, 2005; Melles *et al.*, 2007). It requires far less amount of DNA because multiple bands are derived from across the entire genome. No prior knowledge of DNA sequence is required for AFLP analysis (Vos *et al.*, 1995; Mueller & Wolfenbarger, 1999). In addition, the reproducibility of AFLP is similar to that of other DNA banding pattern-based methods. Finally, AFLP assays are cost-effective and can be automated (Duim *et al.*, 1999) (Applied Biosystems).

The disadvantages of AFLP include the fact that automated analysis equipment may be required due to the many fragments involved and the huge quantity of information generated. Another disadvantage of AFLP that cannot be solved so easily is that the DNA template should not be a mixture of DNAs from various organisms, which prevents the use of AFLP on DNA taken directly from human specimens and environmental samples (Mortimer & Arnold, 2001).

DNA sequencing-based methods (Fig. 2)

DNA sequence, the nucleotide order in a genomic fragment, is the original genetic information of an organism and can be used directly for differentiation and phylogenetic analysis of bacterial strains. The most significant advantages of DNA sequencing-based genotyping over DNA banding pattern-based methods is its high reproducibility because it relies on unambiguous DNA sequences that can easily be stored in online databases and compared among laboratories. GenBank, the largest DNA sequence database, stores huge quantities of genomic sequences as well as locus-specific sequences from almost all known bacteria, and is the most frequently used sequence database by molecular microbiologists. DNA sequencing technology has profoundly influenced microbiology (Salser, 1974; Hall, 2007). DNA sequencing-based genotyping of bacteria has significantly contributed to many aspects of genotyping by identifying SNPs, sequence deletions or insertions (including sequence duplications, such as VNTRs), and genes under positive selection.

Various types of differences among DNA sequences may be identified. SNPs are DNA sequence variations that occur when a single nucleotide – A, T, C, or G – differs between members of a given species. This is the most common type of genetic variation (Schork *et al.*, 2000). In addition, deletion, insertion, and duplication events occur widely in bacterial genomes and are regarded as important evolutionary mechanisms. These changes carry the same genotypic weight as SNPs (Lupski & Weinstock, 1992; van *et al.*, 1998; Ogata *et al.*, 2000; Drancourt *et al.*, 2004). DNA sequencing is an efficient way to identify these genetic changes in order to differentiate and phylogenetically classify bacterial strains. Compared with other SNP genotyping methods, sequencing is particularly suited for identification of multiple SNPs within a small region of DNA. In addition to detecting with

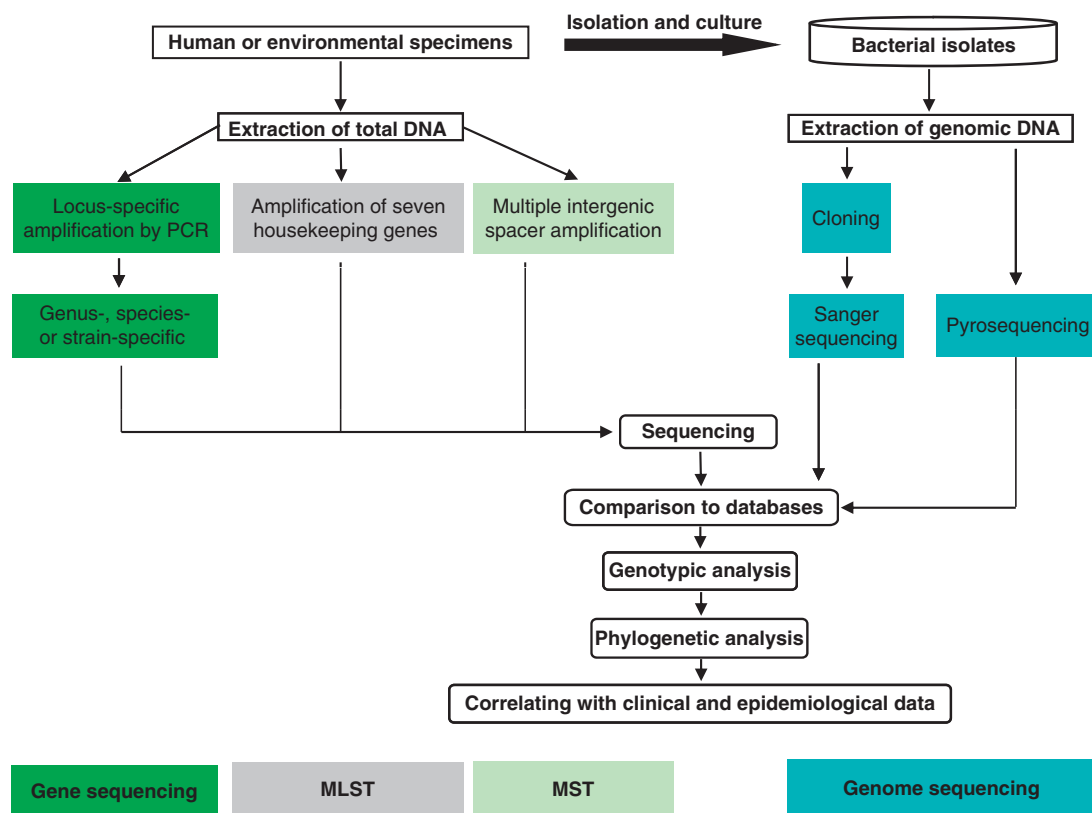


Fig. 2. Flow chart of DNA sequence-based genotyping methods for bacterial strains.

a high sensitivity and specificity sequence differences, DNA sequencing may also allow the evaluation of the evolutionary forces that led to these differences. Relying on the neutral theory of molecular evolution, a high ratio of nonsynonymous over synonymous changes inferred from original DNA sequences can indicate positive selection, which confers a fitness benefit and will thus increase the frequency of that change (Kimura, 1977). A well-known example of positive selection in action is the development of antibiotic resistance in bacteria. When exposed to antibiotics, most bacteria die quickly, but some may experience mutations that make them less susceptible to treatment. Positive selection acts on any mutation that contributes to antibiotic resistance and increases the frequency of those mutations (Aminov & Mackie, 2007). In addition, DNA sequencing-based identification of genes under positive selection is also an effective approach to identify virulence in bacteria (Chen *et al.*, 2006).

Two main strategies in DNA sequencing

Currently, there are two main strategies for DNA sequencing: the traditional Sanger method and the newly developed pyrosequencing method.

The Sanger method

The Sanger method is also referred to as dideoxy sequencing or chain termination DNA sequencing. In this method, the DNA sequence of a single-stranded template DNA is determined using DNA polymerase to synthesize a set of polynucleotide fragments of different lengths through the use of dideoxynucleotides that interrupt the elongation step of DNA amplification (Sanger *et al.*, 1977). Since its description in 1974, the Sanger method has become the most widely used DNA sequencing method. However, recently, new DNA sequencing methods offering much higher throughput at lower cost have revolutionized DNA sequencing. Among these developments, pyrosequencing is the most remarkable.

By comparison with pyrosequencing, Sanger sequencing is more expensive and requires more DNA quantity (several micrograms) and a cloning step for large sequencing projects such as genome sequencing.

Pyrosequencing

In contrast to the conventional Sanger method (Sanger *et al.*, 1977), pyrosequencing is a nonelectrophoretic sequencing method based on real-time quantitative detection of pyrophosphate released following nucleotide incorporation

into a growing DNA chain during DNA synthesis (Nyren, 1987; Ronaghi *et al.*, 1998). The current state of pyrosequencing technology leads to an *c.* 100-fold increase in throughput and a 10-fold reduction in cost over the conventional Sanger method. In 2004, the Roche (454) company first commercialized a high throughput sequencer using the pyrosequencing technology. Following shearing, genomic DNA fragments are hybridized to the surface of agarose beads, and then amplified by emulsion PCR and sequenced, without any cloning step (Margulies *et al.*, 2005). The current output of the Roche instrument is *c.* 100 Mb of sequence data per 7-h run, with an average read length of *c.* 250 bp. More recently, Illumina (Solexa) developed a pyrosequencer that uses glass-attached oligonucleotides that are complementary to specific adapters previously ligated onto DNA library fragments. Sequences are obtained by amplification using an isothermal polymerase, and Sanger-like sequencing. The Illumina method produces *c.* 40–50 million sequence reads of 32–40 bp simultaneously. The latest pyrosequencing competitor is Applied Biosystems. Their SOLID instrument uses a sequencing process catalyzed by DNA ligase. Oligonucleotide adaptor-linked DNA fragments are coupled with magnetic beads that are covered with complementary oligonucleotides. Each bead-bound DNA fragment is amplified using emulsion PCR, and sequencing is obtained using a thermostable ligase. The SOLID instrument produces 3–4 Gb of sequence data with an average read length of 25–35 bp.

To date, pyrosequencing has been used for several applications of genotyping. Sequencing the highly polymorphic regions in 16S rRNA gene by pyrosequencing demonstrated to be a rapid and inexpensive approach for identification and differentiation of bacteria (Jonasson *et al.*, 2002, 2007). In a recent article, Dethlefsen *et al.* (2008) identified 5700 different bacterial taxa in stool specimens from patients before and after antibiotic therapy. Detection of antibiotic resistance has also benefited from pyrosequencing. Mutation detection in antibiotic resistance genes of *M. tuberculosis* by pyrosequencing has been shown to be a cost-effective, highly accurate, and high-throughput method for drug resistance screening (Arnold *et al.*, 2005; Jureen *et al.*, 2006). Grouping, typing, and subtyping bacteria with specific virulence or of epidemiological interest may also be obtained. Pyrosequencing-based SNP detection in three genes, *siaD*, *porB*, and *porA*, allow a rapid and specific characterization of *Neisseria meningitidis* strains (Diggle & Clarke, 2004). Pyrosequencing of genomic fragments has been used to genotype *Neisseria gonorrhoeae* and *B. anthracis* (Diggle & Clarke, 2004; Unemo *et al.*, 2004; Wahab *et al.*, 2005).

In comparison with Sanger sequencing, pyrosequencing, although being less expensive and having a higher output, is limited by short read lengths (25–250 bp), small numbers of

samples that can be run simultaneously, and difficult sequence assembly. The ability to generate DNA sequences rapidly and the potential for scale-up make pyrosequencing a valuable option for SNP genotyping (Clarke, 2005).

Gene sequencing

16S rRNA gene

16S rRNA gene is highly conserved among bacteria and other kingdoms because rRNA is essential for the survival of all cells due to its involvement in protein synthesis (Woese, 1987; Hillis *et al.*, 1991). Consequently, amplification and sequencing of the 16S rRNA gene is widely used for identification and phylogenetic classification of prokaryotic species, genera, and families (Woese, 1987; Stackebrandt & Goebel, 1994). At the strain level, however, 16S rRNA gene is too conserved to be useful, with the exception of few species such as *N. meningitidis* (Sacchi *et al.*, 2002), *Mycoplasma agalactiae* and *Mycoplasma bovis* (Konigsson *et al.*, 2002). In practice, 16S rRNA gene sequencing is mostly used for bacterial species identification.

Other genes

Some less-conserved genes, especially those under positive selection, have often been used for species identification and, in some rarer cases, for bacterial strain typing. Most of these genes encode bacterial surface proteins or virulence factors. The choice of appropriate genes may vary according to the species.

Encoding the β subunit of RNA polymerase, *rpoB* appears to exist in only one copy in most bacteria and has widely been used for bacterial species identification and subtyping (Mollet *et al.*, 1997; Adekambi *et al.*, 2006). The *rpoB* sequence offers higher discriminatory power than 16S rRNA gene (De & De, 2004). DNA sequencing of the variable regions in the *slpA* gene, which encodes a surface layer protein of *C. difficile*, were found to be an ideal alternative to serotyping because the sequence variations are identical within a given serogroup but divergent between serogroups (Karjalainen *et al.*, 2002; Kato *et al.*, 2005). The *set* genes (*set2*, *set5*, and *set7*), which encode *S. aureus* superantigen-like proteins, were found to exhibit a higher discriminatory power than MLST based on seven genes (Aguiar-Alves *et al.*, 2006). Sequence analysis of the *ompA* gene, which encodes a major antigenic outer membrane protein of *Rickettsia* species, was demonstrated to be an important tool for identifying and subtyping these bacteria (Fournier *et al.*, 1998, 2003). Sequencing of the X region of the protein A gene (*spa*) is widely used for subtyping methicillin-resistant *S. aureus* (MRSA) strains and proved to be nearly as discriminatory as PFGE (Shopsin *et al.*, 1999; Koreen *et al.*, 2004; Deurenberg *et al.*, 2007). For more informative

discrimination of MRSA strains, *spa* is strongly suggested to be combined with other gene loci or other genotyping methods (Hallin *et al.*, 2007; Kuhn *et al.*, 2007).

Although genes encoding surface proteins or virulence factors evolve very quickly, interpretation of the typing results may be misleading, especially in evaluating the bacterial population structure or in the case of a long-term epidemiological survey. Combination of genes encoding surface antigens or virulence factors with housekeeping genes may be a rational approach for bacterial strain typing, particularly for bacterial species with a high rate of genetic recombination (Cai *et al.*, 2002).

MLST

MLST is a method using DNA sequencing to uncover allelic variants in several conserved genes (usually seven genes), and is currently one of the most popular genotyping methods for characterizing bacterial strains (Maiden *et al.*, 1998). MLST examines multiple housekeeping genes whose sequences are constrained because of the essential function of the proteins they encode; the variation observed in these sequences is therefore neutral or nearly neutral.

Typically, fragments of 450–500 bp of seven genes are sequenced and each different sequence for a given gene is attributed a number. Each strain is, therefore, assigned a seven-number allelic profile designated as sequence type (ST) (Maiden *et al.*, 1998; Spratt, 1999). MLST is suitable for long-term investigation of bacterial population structures, particularly when subtyping bacterial species with a high rate of genetic recombination, such as *N. meningitidis* (Maiden *et al.*, 1998), *S. pneumoniae* (Enright & Spratt, 1998), and *Enterococcus faecalis* (Ruiz-Garbajosa *et al.*, 2006). MLST analysis indicates that recombinational replacements in many species contribute more to clonal diversification than do point mutations (Feil & Spratt, 2001). DNA sequences are easily stored in online databases, which allow convenient exchange of strain typing data both within and between laboratories. These databases also facilitate the global epidemiological survey of bacterial infections (Feil & Spratt, 2001; Jolley *et al.*, 2004; Aanensen & Spratt, 2005). So far, MLST has been applied to more than 23 bacterial species and is regarded as a reference genotyping method for many bacteria. Detailed information can be found on the MLST homepage, <http://www.mlst.net> (Table 2).

However, MLST also has drawbacks. First, alleles are assigned to a numbering system that is not representative of the actual gene sequence, which makes the phylogenetic analysis of tested strains poorly credible (Clarke, 2002). Second, the use of highly conserved housekeeping genes in MLST often fails to detect the variability of closely related strains. Finally, sequencing of seven genes is costly and time consuming.

Sequencing of noncoding DNA

16S–23S rRNA gene internal transcribed spacer (ITS)

ITS, a special genomic region separating 16S and 23S rRNA genes in prokaryotic microorganisms, consists mainly of noncoding sequences and is the first noncoding sequence used for bacterial strain typing (Gurtler & Stanisich, 1996; Sadeghifard *et al.*, 2006). ITS varies not only in sequence and length but also in the number of alleles and their positions on the chromosome (Garcia-Martinez *et al.*, 1999; Sadeghifard *et al.*, 2006). The high polymorphism of the ITS locus makes it useful for identifying and subtyping bacteria (Gurtler & Mayall, 1999). While size variation of ITS resulting from either the amplification or digestion of amplicons by REs is useful for bacterial strain typing, sequencing ITS can provide a more comprehensive analysis of polymorphisms for strain typing. However, more than one copy of ITS is often found in the bacterial genome, and those ITS sequences within a given strain may differ from one another, making direct DNA sequencing of ITS in some bacterial species difficult. Cloning-based sequencing can resolve this problem but makes it a time-consuming method (Garcia-Martinez *et al.*, 1999).

One advantage of using ITS for studying phylogeny, molecular evolution, or population genetics is that ITS can be amplified easily by picking primers in the conserved region of 16S rRNA and 23S rRNA genes (Gurtler & Stanisich, 1996; Gurtler & Mayall, 1999). In comparison with 16S rRNA gene, ITS is more variable and exhibits greater resolution for subtyping bacteria at the strain level. For the *Bartonella* genus, partial ITS amplification and sequencing has proven to be a sensitive tool for differentiating *B. henselae*, *Bartonella clarridgeiae*, and *Bartonella bacilliformis* isolates (Birtles *et al.*, 2000; Houpijian & Raoult, 2001). However, ITS is mainly used for species or subspecies identification, and less for strain typing. In addition, it is not suitable for bacteria from the order *Rickettsiales* because their 16S and 23S rRNA genes are not contiguous.

Multispacer typing (MST)

MST, a novel genotyping system first applied to *Y. pestis* in 2004, is a DNA sequencing-based genotyping method that uses as typing markers highly variable intergenic spacers (Drancourt *et al.*, 2004). MST assumes that noncoding DNA sequences, which are subject to less selection pressure than genes, vary more than genes and so are preferable for bacterial strain typing (Drancourt *et al.*, 2004). The intergenic spacers used as typing markers in MST are chosen to be those noncoding sequences varying the most between aligned genomes of bacterial strains within one species, or between closely related species if only one genome sequence is available for a given species

(Drancourt *et al.*, 2004; Fournier *et al.*, 2004; Li *et al.*, 2006b). For example, the genome sequences of the closely related *B. henselae* and *Bartonella quintana* were compared to select the most variable intergenic spacers, based on the hypothesis that the intergenic spacers varying the most between closely related species would also vary the most among all strains of the species (Li *et al.*, 2006b). MST based on six highly variable intergenic spacers identified 19 MST types among 36 *Y. pestis* strains from three biovars detected in dental pulp from patients deceased from plague in the second and third pandemics (Drancourt *et al.*, 2004). This work demonstrated the great potential of MST for bacterial strain typing.

With rapid accumulation of bacterial genome sequences and the increasing number of species with more than two genome sequences available, the highly variable intergenic spacers selected by genome comparison are becoming more representative and powerful for bacterial strain typing. After amplification and sequencing, any DNA sequence variation in a spacer can provide a unique ST. Combination of STs from each studied spacer provides an MST genotype. The phylogenetic organization of studied strains, however, was inferred from concatenation of all selected intergenic spacers (Fig. 3).

MST has been applied successfully to several human pathogens, including *Y. pestis* (Drancourt *et al.*, 2004), *Rickettsia conorii* (Fournier *et al.*, 2004), *Rickettsia prowazekii* (Zhu *et al.*, 2005), *Rickettsia sibirica* (Fournier *et al.*, 2006), *Coxiella burnetii* (Glazunova *et al.*, 2005), *B. henselae* (Li *et al.*, 2006b, 2007), *B. quintana* (Foucault *et al.*, 2005) and *Tropheryma whippelii* (Li *et al.*, 2008). Comparison of MST and MLST showed the former to have higher resolution (Fournier *et al.*, 2004; Li *et al.*, 2006b). Because MST is a PCR-based technique, it has the potential to be used directly on noncultured samples. Indeed, MST has been successful in subtyping *Y. pestis* from dental pulps *c.* 1500 years old (Drancourt *et al.*, 2004). Recently, MST was also used to evaluate the genetic diversity of *B. henselae* and *R. conorii* directly from human specimens (Li *et al.*, 2007). Another advantage of MST is that primers can be chosen in conserved regions of the flanking genes, making amplification of highly variable spacers easy.

While intergenic spacers are considered as rapidly evolving markers, MST has succeeded in establishing reliable correlations of MST genotypes with geographic distribution, clinical manifestations, and epidemiology of strains (Drancourt *et al.*, 2004; Fournier *et al.*, 2004; Zhu *et al.*, 2005; Arvand & Viezens, 2007; Li *et al.*, 2006b, 2007). In addition, MST enabled

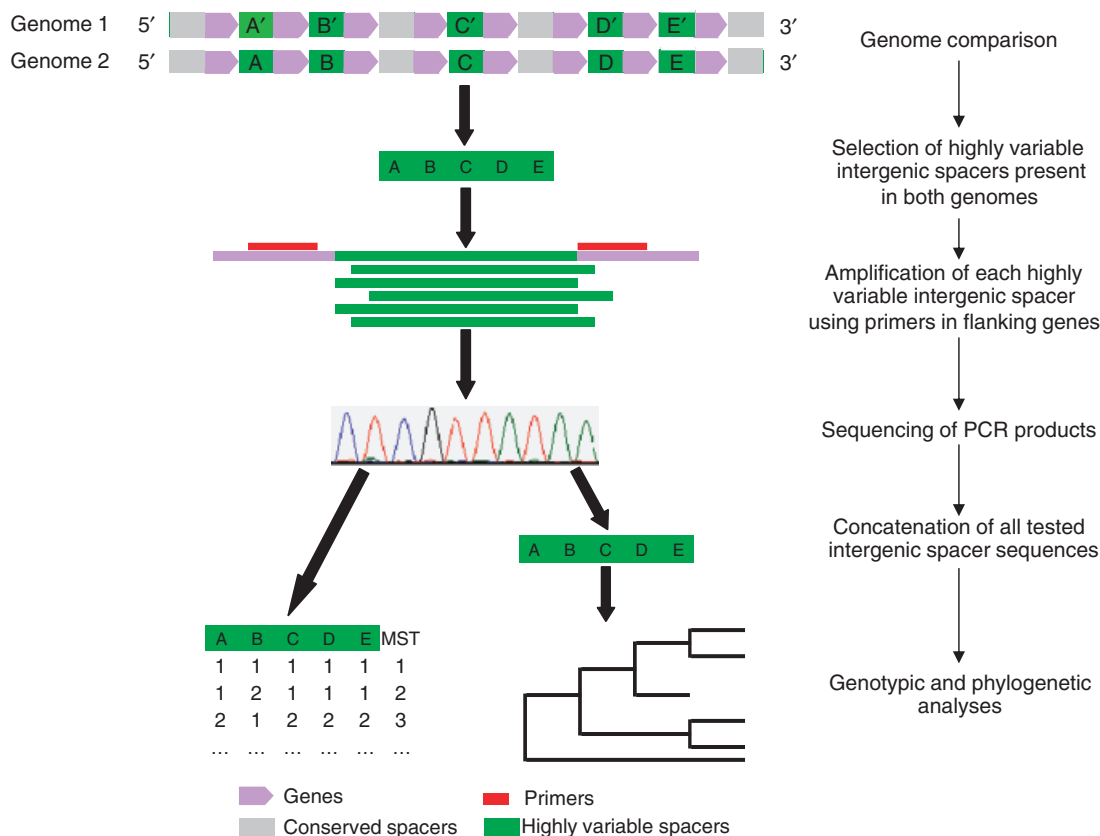


Fig. 3. Outline of MST.

retrospective analysis of *Y. pestis* and *R. prowazekii* in ancient human remains (Drancourt *et al.*, 2004; Zhu *et al.*, 2005). A recent study applying MST to *B. henselae* detected directly from human specimens offers further insight into the genetic diversity of this complex species and suggests that MST is currently the most suitable genotyping tool for evaluating the population structure of *B. henselae* (Li *et al.*, 2007). An online database, MST-Rick, which provides a local BLAST program, enables anyone to compare his own spacer sequences and determine MST genotypes (Table 2).

Genome sequencing

By giving access to the complete genetic content of bacterial strains, genome sequencing is the ultimate genotyping method. Both the Sanger and pyrosequencing methods are currently used for genome sequencing, either separately or combined. The former technology has been the mostly used worldwide since the genome of *Haemophilus influenzae* was sequenced in 1995 (Fleischmann *et al.*, 1995). Genome sequencing using the Sanger method requires the construction of a cDNA library and produces 650–800-bp reads, with a maximum output of 0.44 Mb per 7-h run (Fleischmann *et al.*, 1995; Binnewies *et al.*, 2006). In contrast, pyrosequencing produces short reads (25–250 bp) but with an output of up to 3–4 Gb per run.

Because of the short read length, making sequence assembly difficult, pyrosequencing may currently be preferred for genome resequencing than for *de novo* assembly, although a combination of Sanger sequencing and next generation sequencing has been proposed as a valuable alternative for this latter purpose. Recently, pyrosequencing was used for rapid genome sequencing of *F. tularensis*, a potential bioterrorism bacterial species, in order to identify the detailed genetic differences between a clinically isolated strain and other *F. tularensis* strains (La *et al.*, 2008). Pyrosequencing also proved to be useful for metagenomic studies (Eckburg *et al.*, 2005).

However, although sequencing techniques have been greatly improved since the first bacterial genome was sequenced in 1995 (Fleischmann *et al.*, 1995), with the current possibility of obtaining a genome sequence from a single bacterial cell (Ishoey *et al.*, 2008), genome sequencing remains as yet too expensive and time demanding to be used in routine genotyping. In contrast, both non-sequence-based and sequence-based genotyping methods may benefit from the study of genomic sequences, such as '*in silico*' design of macrorestriction profiles for enzymes used in PFGE, RFLP analyses or microarrays, identification of SNPs, tandem repeats (VNTRs), and variable intergenic spacers (MST).

DNA hybridization-based methods (Fig. 4)

DNA hybridization is widely used for detecting DNA mutations because it requires sequence complementarity to occur. Two main elements, probes and targets, are involved in DNA hybridization. Probes are DNA fragments of known sequences, whereas targets are the free nucleic acids whose identity and abundance are detected by hybridization with fluorescently labeled probes based on their complementarity to the probes.

A DNA array enables genomic DNA to be tested for its ability to hybridize to hundreds or tens of thousands of DNA fragments or oligonucleotides (spots) arrayed on a substrate. Substrates can be membranes, glass, plastic, silicon wafers, or metal alloys. An array is a very powerful tool for studying the transcriptome as well as the genetic diversity of bacteria. Two classes of DNA arrays, macroarrays and microarrays, that differ in the size and number of spots on the supports are used. Macroarrays are usually printed on nylon membranes *c.* 8 × 12 cm in size and contain up to 5000 spots, each > 300 µm in diameter. A microarray, in contrast, is much denser than a macroarray, with up to 1 000 000 spots per array.

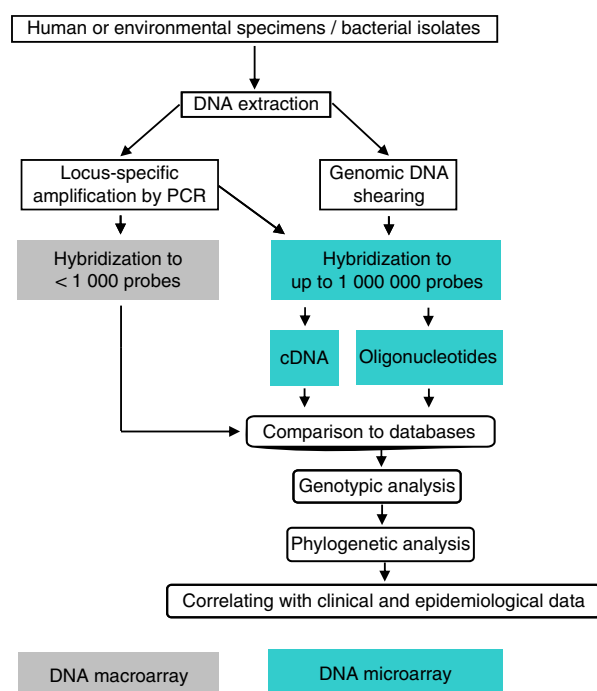


Fig. 4. Flow chart of DNA hybridization-based genotyping methods for bacterial strains.

equipment, and has proven particularly effective for detecting genes involved in antibiotic resistance (Johansen *et al.*, 2003; Zhang *et al.*, 2007). In a recent study, one group rapidly detected mutations in *rpoB*, *katG*, *inhA*, and *ahpC*, which are associated with rifampin and isoniazid resistance in *M. tuberculosis*. The macroarray approach in this study showed high sensitivity and specificity and could even be directly applied to DNA extracted from clinical specimens (Zhang *et al.*, 2007).

Spoligotyping, detection of the polymorphism in the direct variant repeats (DVRs) loci of the mycobacterial chromosome, is a PCR-based macroarray technique for simultaneous detection and differentiation of MTC bacteria (Groenen *et al.*, 1993). The DVR loci, comprising a cluster of well-conserved 36-bp direct repeats (DRs) and spacers of 35–41 bp interspersed throughout DRs (van Embden *et al.*, 2000), show considerable strain-to-strain polymorphism (Fang *et al.*, 1998). Amplification of the DVRs using primers complementary to the DRs, followed by the hybridization of amplicons to the oligonucleotides derived from the spacer sequences (probes), generates distinct spoligotyping patterns. Spoligotyping has been proven to be a rapid, reproducible, informative, and effective technique for use on DNA taken directly from human specimens or environmental samples without the need for prior culture. It has been widely used in epidemiological studies of *M. tuberculosis* infections, especially in low-income countries (Goyal *et al.*, 1997; Hayward & Watson, 1998; Heyderman *et al.*, 1998). An international spoligotyping database, SpolDB4, suggested the existence of geographical genetic clones within MTC populations and provided a large-scale conceptual framework for the global TB Epidemiologic Network (Brudey *et al.*, 2006) (Table 2). For *Corynebacterium diphtheriae*, a macroarray-based spoligotyping method based on hybridization analysis of two CRISPR loci differentiated 20 strains into three spoligotypes (Mokrousov *et al.*, 2008). Eighty *S. aureus* isolates were classified within 52 genotypes using a DNA macroarray containing 465 intragenic amplicons (Trad *et al.*, 2004).

By comparison with microarray, DNA macroarray is less expensive but suffers from a lower discriminatory power due to the small number of characters studied.

DNA microarrays

DNA microarrays are specially treated microscope slides (chips) that carry an ordered mosaic of sequences representing most or all of the genes of an organism. Since its development in 1995 (Schena *et al.*, 1995), the DNA microarray technique has quickly become an efficient tool for investigating the bacterial transcriptome (DeRisi *et al.*, 2000; Khodursky *et al.*, 2000) and genetic diversity (Hacia, 1999; Garaizar *et al.*, 2006). In contrast to Southern blotting and

macroarray techniques, where a limited number of probes are transferred onto nitrocellulose or nylon membranes, a DNA microarray includes tens of thousands of probes arrayed directly on the support and it therefore constitutes a high-throughput genotyping method (Garaizar *et al.*, 2006). However, compared with other locus-limited genotyping methods, the DNA microarray remains more expensive. In addition, satisfactory quality controls are difficult to achieve in DNA microarray analysis because many factors affect nucleic acid hybridization reactions. The probe design in a DNA microarray is mainly based on reference DNA sequences, including prior knowledge of the polymorphisms derived from other genotyping studies. However, comparison of multiple genomes within a given bacterial species reveals numerous 'accessory genes' (Fraser-Liggett, 2005; Tettelin *et al.*, 2005; Lefebvre & Stanhope, 2007), some of which may be absent in the reference genome. As a result, DNA microarray analysis may underestimate the genetic diversity among the bacterial population.

There are two kinds of DNA microarrays: cDNA microarray and oligonucleotide microarray, which differ in the size of the probes used. A cDNA microarray uses cDNA as probes and is often used to identify the presence or absence of genes in a technique called comparative genomic hybridization. An oligonucleotide microarray, on the other hand, uses short oligonucleotides as probes and is usually used to identify SNPs.

cDNA microarrays

The probes in a cDNA microarray are complete genes amplified by PCR or taken from cDNA libraries (Duggan *et al.*, 1999). A whole-genome microarray is a powerful tool for investigating comprehensive variation at the genomic level (Dorrell *et al.*, 2001; Garaizar *et al.*, 2006). In addition, cDNA microarrays are useful for identifying genes associated with serotypes, virulence, antibiotics resistance, epidemics, and host diversity. These types of genes are difficult to identify by most other genotyping methods (Garaizar *et al.*, 2002; Borucki *et al.*, 2004). Once these genes are identified by microarray analysis, they may serve as probes for individual discrimination of unknown bacteria. In a powerful illustration of cDNA microarrays, an array including a series of housekeeping genes, virulence genes, and antibiotic resistance-associated genes was used for the simultaneous identification, characterization, and differentiation of *S. aureus*, *E. coli*, and *Pseudomonas aeruginosa* (Cleven *et al.*, 2006). Specific loci-based cDNA microarrays are cheaper, easier to use, and give results that are easier to interpret compared with whole-genome cDNA microarrays.

Because cDNAs or partial cDNAs were used as probes in the cDNA microarray and the hybridization is often under conditions of lower stringency, a family of genes that share

significant sequence identity (> 70%) may cross-hybridize with the probes, leading to unspecific hybridization.

Oligonucleotide microarrays

The probes of an oligonucleotide microarray are short nucleic acid sequences with *c.* 30–70 nucleotides spotted or directly synthesized in high density on the solid substrate (up to 1 000 000 oligonucleotides per chip) (Lockhart *et al.*, 1996; Iwasaki *et al.*, 2002). Hybridization on an oligonucleotide microarray is more stringent than on a cDNA microarray, which reduces cross-hybridization. In addition, in contrast to the cDNA microarray, which can only indicate the presence or absence of genes, an oligonucleotide microarray can detect SNPs and short sequence deletions by dividing a gene sequence into a number of oligonucleotide probes (Lockhart *et al.*, 1996; Kato-Maeda *et al.*, 2001; Iwasaki *et al.*, 2002). All of the possibilities at the polymorphic sites within an oligonucleotide can be deduced from the results of an oligonucleotide microarray. In fact, well-designed oligonucleotide microarrays using short nucleotides (10–20 mers) as probes can provide investigators with actual sequence data. These are termed resequencing microarrays (Hacia, 1999; Dunman *et al.*, 2004; Zwick *et al.*, 2005). DNA resequencing helps determine the order of nucleotides in a DNA fragment by hybridization with the genomic reference sequence (Zhan & Kulp, 2005). Oligonucleotide microarray-based resequencing is a high-throughput method for SNP detection and can even be used for high quality whole-genome resequencing (Zwick *et al.*, 2005). The whole-genome resequencing microarray has been shown to be a rapid, highly discriminatory, and cost-effective method for strain typing compared with the Sanger method of sequencing whole genomes (Mockler *et al.*, 2005; Zwick *et al.*, 2005).

Locus-specific oligonucleotide microarrays are also used for genotyping. The complexity of *E. coli* serotypes makes antibody-based typing very difficult. Oligonucleotide microarrays containing the *E. coli* genes for O- and H-antigens are valuable alternatives to classical serotyping because they allow single colonies to be processed within one working day (Liu & Fratamico, 2006; Ballmer *et al.*, 2007).

Oligonucleotide microarrays are more discriminatory than a cDNA microarray due to the use of short oligonucleotides as probes. A cDNA microarray, however, can be used with environmental samples or human specimens if the cDNA is amplified beforehand (Zhou, 2003; Roth *et al.*, 2004).

Selection of appropriate genotyping methods

An ideal genotyping method would have the following characteristics: applicable to all isolates (validity), capable of differentiating unrelated isolates (discriminability),

reproducible both within and between laboratories (reproducibility), rapid, cost efficient, and easy to perform. However, among the many genotyping methods that are currently available, no single method is universally ideal because each one has both strengths and weaknesses, as discussed above.

Two main objectives of genotyping may be studied using distinct methods. Surveillance of local outbreaks of bacterial infections may be obtained using fast-evolving markers such as those studied by MST (Fournier *et al.*, 2004; Li *et al.*, 2006b) or MLVA (van *et al.*, 1998). In contrast, for long-term epidemiological studies, or population studies, conserved and stable marker-based typing methods such as MLST, PFGE, and whole-genome sequencing may be more appropriate (Cooper & Feil, 2004) (Fig. 5).

Discriminatory power, or resolution, which refers to the ability to distinguish bacterial strains from a given species, is one of the most important criteria for selecting a genotyping method. For example, bacterial populations may appear to be identical using phenotypic or low-resolution genotyping analysis, but different using high-resolution genotyping methods. Generally, methods studying fast-evolving markers, such as MST, MLVA, DGE, or HRM are more discriminatory than those studying slow-evolving markers such as MLST. Whole genome-based typing methods such as PFGE, AFLP, RFLP, and genome microarray are more discriminatory than locus-limited typing methods such as ribotyping and PCR-RFLP. However, the discriminatory power is also influenced by many factors, such as the enzymes and primers used and the enzymatic and amplification conditions. In addition, most genotyping methods are strain sensitive, and their discriminatory power may vary according to the species studied.

Reproducibility, the ability to obtain the same result whenever the same test is performed, is also important both within and between laboratories, to allow global epidemiological surveillance of bacterial infections. In addition to methodological reproducibility, the stability of typing methods should also be considered when choosing a typing method. Generally, results from banding pattern-based methods are less reproducible and comparable than sequence-based methods.

Another criterion that should be considered is the purity of the strain(s) to be studied. Given that > 99% of environmental bacteria cannot be cultured, PCR-based genotyping methods, namely PCR-RFLP, MLVA, DGE, HRM, MLST, MST, and cDNA microarrays, are better suited than other methods for polymicrobial environmental or human samples (Kellenberger, 2001; Rappe & Giovannoni, 2003) (Fig. 5). In contrast, although they are PCR-based, AFLP, RAPD, and REP-PCR, using nonspecific primers, are poorly applicable to polymicrobial samples. PFGE, RFLP, oligonucleotide microarrays, and genome sequencing,

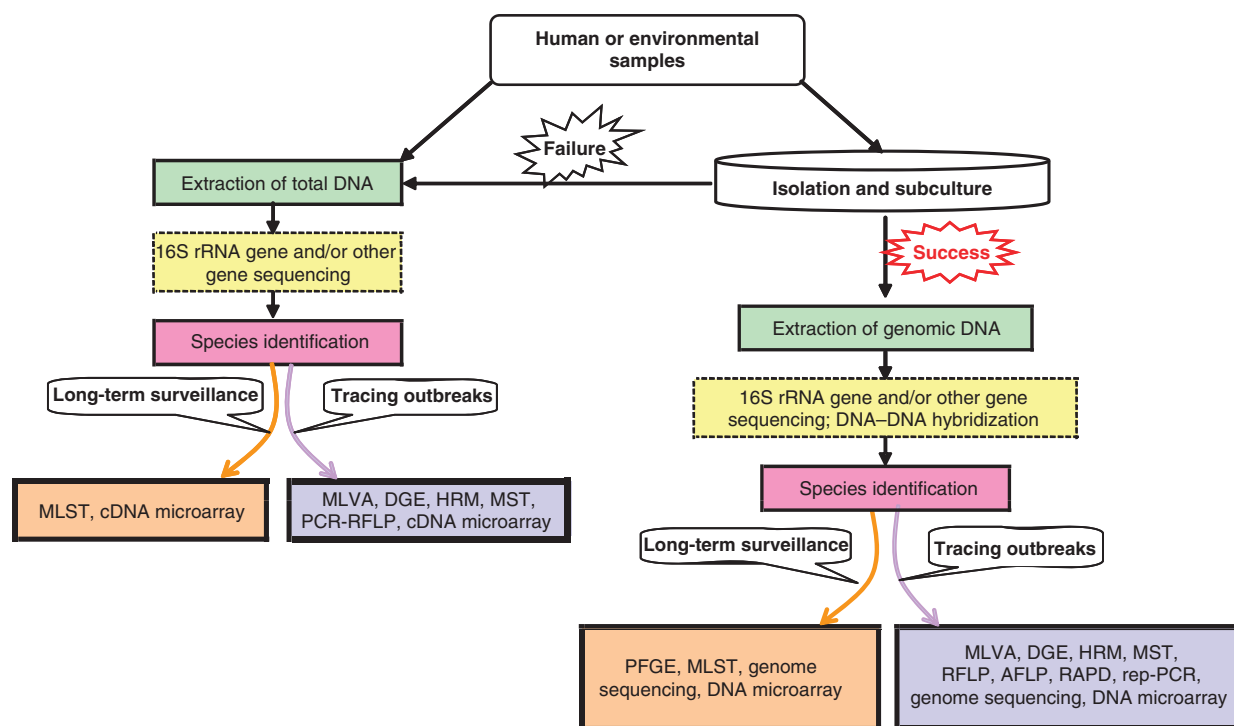


Fig. 5. Selection of appropriate genotyping method for subtyping bacterial strains.

require large amounts of high-quality genomic DNA, which means that they are not suited for typing bacteria in polymicrobial samples as well (Fig. 5).

Other important considerations when choosing genotyping methods include the cost of reagents, equipment, and staff; the time necessary to obtain results; and the difficulty of the technique. Generally, in comparison with unambiguous DNA sequencing-based methods, which require an expensive DNA sequencer, methods based on DNA banding pattern are less expensive but are also less discriminatory and reproducible. The DNA microarray is more likely to produce high-throughput strain typing results, but it is also more expensive than banding pattern- and sequencing-based methods.

Current challenges in bacterial strain typing

In molecular biology, a bacterial species is defined as a relatively homogeneous population sharing at least 70% hybridization of DNA (Wayne *et al.*, 1987; International Committee on Systematic Bacteriology, 1988), and over 97% 16S rRNA gene sequence similarity (Stackebrandt & Goebel, 1994). Over the last decade, as a result of the increasing number of bacterial genome sequences available, comparative genomics demonstrated that genetic diversity within bacterial species was far greater than previously

thought (Fraser-Liggett, 2005; Tettelin *et al.*, 2005; Binnewies *et al.*, 2006; Lefebvre & Stanhope, 2007). This also introduced the concept of bacterial pan-genome, including the 'core genome,' made of genes shared by all strains in a species, and the 'dispensable' or 'flexible genome' comprising genes unique in one or some strains. Two strains of the same bacterial species may differ in gene content by as much as 30% and a species may show endless diversity (Fraser-Liggett, 2005; Tettelin *et al.*, 2005). Genomic variations may also involve sequences from genes shared by the majority of strains within a species, with possibility of intragenic variable and conserved fragments. The concept of pan-genome highlights a great challenge of bacterial strain typing, that is, how can a strain typing method evaluate the genetic diversity of a bacterial species? As an example, the genomic diversity of *Streptococcus agalactiae* based on whole genome analysis was found to be inconsistent with MLST sequence types (Medini *et al.*, 2005). This discrepancy may be explained by the fact that housekeeping genes used in MLST are selected from the core genome rather than the pan-genome. Because intraspecies diversity results from point mutations, insertions/deletions, and genetic recombination, the recombination rate should be taken into account when designing a typing procedure for a given species. Some bacteria, such as streptococci, meningococci, *Helicobacter pylori*, and salmonellae are highly dynamic, with high rates of genetic recombination. Therefore, typing methods

studying few genetic markers may not be sufficient for genotyping these bacteria at the strain level (Medini *et al.*, 2005), and comparison of genomic sequences from several strains may facilitate selection of appropriate targets for a given species.

Another challenge in microbiology is species identification. Species identification may be regarded as bacterial typing at the species level. Some strain genotyping approaches based on relatively conserved genetic markers such as MLST, ribotyping, 16S rRNA gene-/ITS-sequence typing have been used both for species identification and strain typing. In contrast, fast-evolving markers such as VNTRs, noncoding DNA, and horizontally transferred genes, are not suitable for species identification and may misclassify strains in a wrong species. Polymicrobial specimens, especially those containing unculturable strains, are especially problematic for species determination. Metagenomic studies, notably those using highly conserved genes such as 16S rRNA gene, may be particularly suited for bacterial strain identification in these samples (Fig. 5). After species identification, appropriate strain typing methods may be applied to studied strains.

Unraveling the association between genotypes and phenotypic or epidemiological traits such as bacterial virulence, antibiotic resistance, host adaptation, geographic origin, pandemic, or epidemic outbreaks, is another major challenge of bacterial strain genotyping. Although genotyping is highly discriminatory, it is still important to take epidemiological traits into account when evaluating putative outbreaks of bacterial infections. Otherwise, a false outbreak may be announced. While most typing markers are not directly virulence or antibiotic resistance genes, it is possible to find correlations between specific genotypes and phenotypic or epidemiological traits by analyzing the genetic distances between strains through phylogeny. In addition, the inclusion of a large collection of strains and different typing markers is also helpful for linking genotypes with phenotypical and epidemiological traits (Caws *et al.*, 2008). While phylogenetic profiling is useful for correlating phenotype to genotype as well as for the annotation and identification of functionally related genes, this informative computerized method is currently based on genome sequence analysis (Antonov & Mewes, 2008; Benfey & Mitchell-Olds, 2008). With increasing accumulation of bacterial genotypes generated by various approaches, application of powerful phylogenetic profiling for linking genotypes identified by strain typing with phenotypical and epidemiological traits is a great challenge in the genomic era.

Finally, will whole genome sequencing replace other genotyping tools? Whole genome sequencing provides the full genetic information of studied strains and thus appears as an ideal strain-typing method. Currently, the limitations of whole genome sequencing for genotyping include the cost, require-

ment for large amounts of high-quality genomic DNA, and the small read length of high throughput sequencing methods. However, due to the continuous reduction of costs, wider application of pyrosequencing, development of strategies that enable sequencing from a single cell, and anticipation that read length will increase, this strategy may become a first-line genotyping tool in the next future.

Concluding remarks

Bacterial strain-typing methods have been rapidly evolving over recent years. The field has been moving toward computer-assisted design/analysis and increasing automation, resolution, throughput, and reproducibility. In the genomic era, bacterial genotyping benefits increasingly from bioinformatics, such as computer-based selection of typing markers; design of typing strategy; storage, exchange, and comparison of genotyping data (DNA banding patterns, DNA sequences, microarray profiles); development of improved analysis tools; and phylogenetic analysis. Genotyping databases containing banding patterns, DNA sequences, or DNA microarray profiles are convenient for interlaboratory comparison, retrospective studies, and long-term epidemiological surveillance of bacterial infections. Locus-specific typing methods now use multiple loci rather than only one, and whole-genome analysis allowed the selection of typing markers to become gradually less empirical and more rational. In addition, newly developed detection systems for discriminating DNA fragments, such as flow cytometry-RFLP, combines the reproducibility and reliability of RFLP with the speed and sensitivity of flow cytometry, which dramatically increases the speed and resolution of conventional genotyping (Larson *et al.*, 2000; Ferris *et al.*, 2004). Other combined methods have recently emerged and have been found to be more functional than the component methods on their own (Chang *et al.*, 2007a).

Typing results obtained from one genotyping method can be used to improve others or even to design novel typing systems. As discussed above, SNP detection is based on prior knowledge of DNA polymorphism generated by other genotyping methods based on DNA sequence. Genome sequencing of one bacterial strain not only provides detailed information of intraspecies diversity but also enables selection of locus-specific typing markers and rational design of genotyping strategies. As examples, MLVA and MST (Drancourt *et al.*, 2004; van, 2007) have benefited from the availability of an increasing number of sequenced bacterial genomes, allowing selection of highly polymorphic VNTR loci or intergenic spacers. Genome sequences also facilitate the rational choice of primers for PCR-based genotyping methods. In addition, genomic sequences are necessary for probe design in DNA microarray-based genotyping (Lucchini *et al.*, 2001; Zhou, 2003).

Bacterial species are relatively heterogeneous entities whose strain diversity is important to understand in order to manage the treatment and epidemiology of infections. Currently, this diversity may be evaluated using genotyping methods classified in three categories: DNA banding pattern-based, DNA sequencing-based, and DNA hybridization-based methods. Each genotyping method offers advantages and disadvantages. Consequently, choosing an appropriate genotyping method is not easy and may depend on the objective studied and several variables including typeability, reproducibility, resolution, cost, ease of execution and interpretation, and time to obtain results. In addition, the results of bacterial genotyping should always be analyzed in the light of clinical and epidemiological data, especially when evaluating novel genotyping methods. Genome sequencing, in addition to enabling the rational selection of adequate targets for many genotyping methods, appears, today, as the ultimate genotyping method, and current technical improvements and reduction of costs and delays may favor this strategy. However, it is difficult to predict whether this method will become the norm and replace other genotyping methods.

References

- Aanensen DM & Spratt BG (2005) The multilocus sequence typing network: mlst.net. *Nucleic Acids Res* **33**: W728–W733.
- Adekambi T, Berger P, Raoult D & Drancourt M (2006) *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *Int J Syst Evol Microbiol* **56**: 133–143.
- Al DS, Tomaso H, Prenger-Berninghoff E, Splettstoesser WD, Scholz HC & Neubauer H (2005) Identification of *Brucella* species and biotypes using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). *Crit Rev Microbiol* **31**: 191–196.
- Aminov RI & Mackie RI (2007) Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol Lett* **271**: 147–161.
- Antonov AV & Mewes HW (2008) Complex phylogenetic profiling reveals fundamental genotype–phenotype associations. *Comput Biol Chem* **32**: 412–416.
- Arnold C, Westland L, Mowat G, Underwood A, Magee J & Gharbia S (2005) Single-nucleotide polymorphism-based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clin Microbiol Infect* **11**: 122–130.
- Arvand M & Vezens J (2007) Evaluation of pulsed-field gel electrophoresis and multi-locus sequence typing for the analysis of clonal relatedness among *Bartonella henselae* isolates. *Int J Med Microbiol* **297**: 255–262.
- Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehrlich R & Hachler H (2007) Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. *J Clin Microbiol* **45**: 370–379.
- Benfey PN & Mitchell-Olds T (2008) From genotype to phenotype: systems biology meets natural variation. *Science* **320**: 495–497.
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573–580.
- Bikandi J, San MR, Rementeria A & Garaizar J (2004) *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* **20**: 798–799.
- Bingen EH, Denamur E & Elion J (1994) Use of ribotyping in epidemiological surveillance of nosocomial outbreaks. *Clin Microbiol Rev* **7**: 311–327.
- Binnewies TT, Motro Y, Hallin PF, Lund O, Dunn D, La T, Hampson DJ, Bellgard M, Wassenaar TM & Ussery DW (2006) Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* **6**: 165–185.
- Birtles RJ, Hazel S, Bown K, Raoult D, Begon M & Bennett M (2000) Subtyping of uncultured bartonellae using sequence comparison of 16 S/23 S rRNA intergenic spacer regions amplified directly from infected blood. *Mol Cell Probes* **14**: 79–87.
- Blakesley RW (1987) Restriction endonuclease: cleavage, ligation, and sensitivity. *Gene Amplif Anal* **5**: 51–102.
- Borucki MK, Kim SH, Call DR, Smole SC & Pagotto F (2004) Selective discrimination of *Listeria monocytogenes* epidemic strains by a mixed-genome DNA microarray compared to discrimination by pulsed-field gel electrophoresis, ribotyping, and multilocus sequence typing. *J Clin Microbiol* **42**: 5270–5276.
- Brudey K, Driscoll JR, Rigouts L *et al.* (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* **6**: 23.
- Busse HJ, Denner EB & Lubitz W (1996) Classification and identification of bacteria: current approaches to an old problem. Overview of methods used in bacterial systematics. *J Biotechnol* **47**: 3–38.
- Cai S, Kabuki DY, Kuaye AY, Cargioli TG, Chung MS, Nielsen R & Wiedmann M (2002) Rational design of DNA sequence-based strategies for subtyping *Listeria monocytogenes*. *J Clin Microbiol* **40**: 3319–3325.
- Cangelosi GA, Freeman RJ, Lewis KN, Livingston-Rosanoff D, Shah KS, Milan SJ & Goldberg SV (2004) Evaluation of a high-throughput repetitive-sequence-based PCR system for DNA fingerprinting of *Mycobacterium tuberculosis* and *Mycobacterium avium* complex strains. *J Clin Microbiol* **42**: 2685–2693.
- Caws M, Thwaites G, Dunstan S *et al.* (2008) The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* **4**: e1000034.
- Chang A, Liew WC, Chuah A, Lim Z, Lin Q & Orban L (2007a) FluMEP: a new genotyping method combining the

- advantages of randomly amplified polymorphic DNA and amplified fragment length polymorphism. *Electrophoresis* **28**: 525–534.
- Chang PL, Hsieh WS, Chiang CL, Tuohy MJ, Hall GS, Procop GW, Chang HT & Ho HT (2007b) The hsp65 gene patterns of less common *Mycobacterium* and *Nocardia* spp. by polymerase chain reaction-restriction fragment length polymorphism analysis with capillary electrophoresis. *Diagn Microb Infect Dis* **58**: 315–323.
- Chansiripornchai N, Ramasoota P, Bangtrakulnonth A, Sasipreeyajan J & Svenson SB (2000) Application of randomly amplified polymorphic DNA (RAPD) analysis for typing avian *Salmonella enterica* subsp. *enterica*. *FEMS Immunol Med Microbiol* **29**: 221–225.
- Chen KW, Lo HJ, Lin YH & Li SY (2005) Comparison of four molecular typing methods to assess genetic relatedness of *Candida albicans* clinical isolates in Taiwan. *J Med Microbiol* **54**: 249–258.
- Chen SL, Hung CS, Xu J et al. (2006) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *P Natl Acad Sci USA* **103**: 5977–5982.
- Chou CH & Wang C (2006) Genetic relatedness between *Listeria monocytogenes* isolates from seafood and humans using PFGE and REP-PCR. *Int J Food Microbiol* **110**: 135–148.
- Chung M, de LH, Matthews P et al. (2000) Molecular typing of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis: comparison of results obtained in a multilaboratory effort using identical protocols and MRSA strains. *Microb Drug Resist* **6**: 189–198.
- Clarke SC (2002) Nucleotide sequence-based typing of bacteria and the impact of automation. *Bioessays* **24**: 858–862.
- Clarke SC (2005) Pyrosequencing: nucleotide sequencing technology with bacterial genotyping applications. *Expert Rev Mol Diagn* **5**: 947–953.
- Cleven BE, Palka-Santini M, Gielen J, Meembor S, Kronke M & Krut O (2006) Identification and characterization of bacterial pathogens causing bloodstream infections by DNA microarray. *J Clin Microbiol* **44**: 2389–2397.
- Cohn DL & O'Brien RJ (1998) The use of restriction fragment length polymorphism (RFLP) analysis for epidemiological studies of tuberculosis in developing countries. *Int J Tuberc Lung Dis* **2**: 16–26.
- Cooper JE & Feil EJ (2004) Multilocus sequence typing – what is resolved? *Trends Microbiol* **12**: 373–377.
- Davis MA, Hancock DD, Besser TE & Call DR (2003) Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of *Escherichia coli* O157:H7. *J Clin Microbiol* **41**: 1843–1849.
- De CE & De VP (2004) Genotypic diversity among *Bacillus licheniformis* strains from various sources. *FEMS Microbiol Lett* **231**: 91–98.
- de Bruijn FJ (1992) Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl Environ Microb* **58**: 2180–2187.
- Denoeud F & Vergnaud G (2004) Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. *BMC Bioinformatics* **5**: 4.
- DeRisi J, van den HB, Marc P, Balzi E, Brown P, Jacq C & Goffeau A (2000) Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. *FEBS Lett* **470**: 156–160.
- Dethlefsen L, Huse S, Sogin ML & Relman DA (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.
- Deurenberg RH, Vink C, Kalenic S, Friedrich AW, Bruggeman CA & Stobberingh EE (2007) The molecular evolution of methicillin-resistant *Staphylococcus aureus*. *Clin Microbiol Infect* **13**: 222–235.
- Diggle MA & Clarke SC (2004) Genotypic characterization of *Neisseria meningitidis* using pyrosequencing. *Mol Biotechnol* **28**: 139–145.
- Dorrell N, Mangan JA, Laing KG et al. (2001) Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res* **11**: 1706–1715.
- Drancourt M, Roux V, Dang LV, Tran-Hung L, Castex D, Chenal-Francisque V, Ogata H, Fournier PE, Crubezy E & Raoult D (2004) Genotyping, orientalis-like *Yersinia pestis*, and plague pandemics. *Emerg Infect Dis* **10**: 1585–1592.
- Duggan DJ, Bittner M, Chen Y, Meltzer P & Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* **21**: 10–14.
- Duim B, Wassenaar TM, Rigtter A & Wagenaar J (1999) High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Appl Environ Microb* **65**: 2369–2375.
- Dunman PM, Mounts W, McAleese F et al. (2004) Uses of *Staphylococcus aureus* GeneChips in genotyping and genetic composition analysis. *J Clin Microbiol* **42**: 4275–4283.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE & Relman DA (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.
- Enright MC & Spratt BG (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* **144**: 3049–3060.
- Fang Z, Morrison N, Watt B, Doig C & Forbes KJ (1998) IS6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J Bacteriol* **180**: 2102–2109.
- Farlow J, Smith KL, Wong J, Abrams M, Lytle M & Keim P (2001) *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *J Clin Microbiol* **39**: 3186–3192.

- Feberwee A, Dijkstra JR, von Banniseht-Wysmuller TE, Gielkens AL & Wagenaar JA (2005) Genotyping of *Mycoplasma gallisepticum* and *M. synoviae* by Amplified Fragment Length Polymorphism (AFLP) analysis and digitalized Random Amplified Polymorphic DNA (RAPD) analysis. *Vet Microbiol* **111**: 125–131.
- Feil EJ & Spratt BG (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**: 561–590.
- Ferris MM, Yan X, Habbersett RC, Shou Y, Lemanski CL, Jett JH, Yoshida TM & Marrone BL (2004) Performance assessment of DNA fragment sizing by high-sensitivity flow cytometry and pulsed-field gel electrophoresis. *J Clin Microbiol* **42**: 1965–1976.
- Field D, Hughes J & Moxon ER (2004) Using the genome to understand pathogenicity. *Meth Mol Biol* **266**: 261–287.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JE, Dougherty BA & Merrick JM (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Fortini D, Ciannamarcuconi A, De SR, Fasanella A, Battisti A, D'Amelio R, Lista F, Cassone A & Carattoli A (2007) Optimization of high-resolution melting analysis for low-cost and rapid screening of allelic variants of *Bacillus anthracis* by multiple-locus variable-number tandem repeat analysis. *Clin Chem* **53**: 1377–1380.
- Foucault C, La SB, Lindroos H, Andersson SG & Raoult D (2005) Multispacer typing technique for sequence-based typing of *Bartonella quintana*. *J Clin Microbiol* **43**: 41–48.
- Fournier PE, Roux V & Raoult D (1998) Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA. *Int J Syst Bacteriol* **48**: 839–849.
- Fournier PE, Dumler JS, Greub G, Zhang J, Wu Y & Raoult D (2003) Gene sequence-based criteria for identification of new *Rickettsia* isolates and description of *Rickettsia heilongjiangensis* sp. nov. *J Clin Microbiol* **41**: 5456–5465.
- Fournier PE, Zhu Y, Ogata H & Raoult D (2004) Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains. *J Clin Microbiol* **42**: 5757–5766.
- Fournier PE, Zhu Y, Yu X & Raoult D (2006) Proposal to create subspecies of *Rickettsia sibirica* and an emended description of *Rickettsia sibirica*. *Ann NY Acad Sci* **1078**: 597–606.
- Fraser-Liggett CM (2005) Insights on biology and evolution from microbial genome sequencing. *Genome Res* **15**: 1603–1610.
- Fromin N, Hamelin J, Tarnawski S, Roesti D, Jourdain-Miserez K, Forestier N, Teyssier-Cuvelles S, Gillet F, Aragno M & Rossi P (2002) Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environ Microbiol* **4**: 634–643.
- Garaizar J, Porwollik S, Echeita A, Rementeria A, Herrera S, Wong RM, Frye J, Usera MA & McClelland M (2002) DNA microarray-based typing of an atypical monophasic *Salmonella enterica* serovar. *J Clin Microbiol* **40**: 2074–2078.
- Garaizar J, Rementeria A & Porwollik S (2006) DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? *FEMS Immunol Med Mic* **47**: 178–189.
- Garcia-Martinez J, Acinas SG, Anton AI & Rodriguez-Valera F (1999) Use of the 16S–23S ribosomal genes spacer region in studies of prokaryotic diversity. *J Microbiol Meth* **36**: 55–64.
- Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyttia-Trees E, Ribot EM & Swaminathan B (2006) PulseNet USA: a five-year update. *Foodborne Pathog Dis* **3**: 9–19.
- Gilson E, Clement JM, Brutlag D & Hofnung M (1984) A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J* **3**: 1417–1421.
- Glazunova O, Roux V, Freylikman O, Sekeyova Z, Fournous G, Tyczka J, Tokarevich N, Kovacava E, Marrie TJ & Raoult D (2005) *Coxiella burnetii* genotyping. *Emerg Infect Dis* **11**: 1211–1217.
- Goyal M, Saunders NA, van Embden JD, Young DB & Shaw RJ (1997) Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. *J Clin Microbiol* **35**: 647–651.
- Grissa I, Bouchon P, Pourcel C & Vergnaud G (2007) On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. *Biochimie* **90**: 660–668.
- Groenen PM, Bunschoten AE, van SD & van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* **10**: 1057–1065.
- Aguiar-Alves F, Medeiros F, Fernandes O, Gudzuki Pereira RM, Perdreau-Remington F & Riley LW (2006) New *Staphylococcus aureus* genotyping method based on exotoxin (set) genes. *J Clin Microbiol* **44**: 2728–2732.
- Gundry CN, Vandersteen JG, Reed GH, Pryor RJ, Chen J & Wittwer CT (2003) Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clin Chem* **49**: 396–406.
- Gurtler V & Mayall BC (1999) rDNA spacer rearrangements and concerted evolution. *Microbiology* **145**: 2–3.
- Gurtler V & Stanisich VA (1996) New approaches to typing and identification of bacteria using the 16S–23S rDNA spacer region. *Microbiology* **142**: 3–16.
- Gzyl A, Augustynowicz E, Mosiej E, Zawadka M, Gniadek G, Nowaczek A & Slusarczyk J (2005) Amplified fragment length polymorphism (AFLP) versus randomly amplified polymorphic DNA (RAPD) as new tools for inter- and intra-species differentiation within *Bordetella*. *J Med Microbiol* **54**: 333–346.
- Hacia JG (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* **21**: 42–47.
- Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* **210**: 1518–1525.
- Hallin M, Deplano A, Denis O, De MR, De RR & Struelens MJ (2007) Validation of pulsed-field gel electrophoresis and spa typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J Clin Microbiol* **45**: 127–133.
- Harvey SP & Minter JM (2005) Ribotyping of *Burkholderia mallei* isolates. *FEMS Immunol Med Mic* **44**: 91–97.

- Hayward AC (1995) Restriction fragment length polymorphism typing of *Mycobacterium tuberculosis*. *Thorax* **50**: 1211–1218.
- Hayward AC & Watson JM (1998) Typing of mycobacteria using spoligotyping. *Thorax* **53**: 329–330.
- Healy M, Huang J, Bittner T *et al.* (2005) Microbial DNA typing by automated repetitive-sequence-based PCR. *J Clin Microbiol* **43**: 199–207.
- Herschleb J, Ananiev G & Schwartz DC (2007) Pulsed-field gel electrophoresis. *Nat Prot* **2**: 677–684.
- Heyderman RS, Goyal M, Roberts P, Ushewokunze S, Zizhou S, Marshall BG, Makombe R, van Embden JD, Mason PR & Shaw RJ (1998) Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping. *Thorax* **53**: 346–350.
- Hillis DM, Moritz C, Porter CA & Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308–310.
- Ho HT, Chang PL, Hung CC & Chang HT (2004) Capillary electrophoretic restriction fragment length polymorphism patterns for the mycobacterial hsp65 gene. *J Clin Microbiol* **42**: 3525–3531.
- Hoffmaster AR, Fitzgerald CC, Ribot E, Mayer LW & Popovic T (2002) Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States. *Emerg Infect Dis* **8**: 1111–1116.
- Houpiikian P & Raoult D (2001) 16S/23S rRNA intergenic spacer regions for phylogenetic analysis, identification, and subtyping of *Bartonella* species. *J Clin Microbiol* **39**: 2768–2778.
- Hulton CS, Higgins CF & Sharp PM (1991) ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* **5**: 825–834.
- Hyttia E, Bjorkroth J, Hielm S & Korkeala H (1999) Characterisation of *Clostridium botulinum* groups I and II by randomly amplified polymorphic DNA analysis and repetitive element sequence-based PCR. *Int J Food Microbiol* **48**: 179–189.
- International Committee on Systematic Bacteriology (1988) Announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *J Appl Bacteriol* **64**: 283–284.
- Ishoev T, Woyke T, Stepanauskas R, Novotny M & Lasken RS (2008) Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* **11**: 198–204.
- Iwasaki H, Ezura Y, Ishida R, Kajita M, Kodaira M, Knight J, Daniel S, Shi M & Emi M (2002) Accuracy of genotyping for single nucleotide polymorphisms by a microarray-based single nucleotide polymorphism typing method involving hybridization of short allele-specific oligonucleotides. *DNA Res* **9**: 59–62.
- Jeffery N, Gasser RB, Steer PA & Noormohammadi AH (2007) Classification of *Mycoplasma synoviae* strains using single-strand conformation polymorphism and high-resolution melting-curve analysis of the *vlhA* gene single-copy region. *Microbiology* **153**: 2679–2688.
- Jin JD, Lee DS, Shin EK, Kim SJ, Jung R & Hahn TW (2006) Molecular typing by random amplification of polymorphic DNA (RAPD) and detection of virulence genes of *Salmonella enterica* subspecies enterica serovar Gallinarum biovar gallinarum. *J Vet Med Sci* **68**: 1321–1326.
- Johansen IS, Lundgren B, Sosnovskaja A & Thomsen VO (2003) Direct detection of multidrug-resistant *Mycobacterium tuberculosis* in clinical specimens in low- and high-incidence countries by line probe assay. *J Clin Microbiol* **41**: 4454–4456.
- Johansson A, Farlow J, Larsson P *et al.* (2004) Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. *J Bacteriol* **186**: 5808–5818.
- Jolley KA, Chan MS & Maiden MC (2004) mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* **5**: 86.
- Jonasson J, Olofsson M & Monstein HJ (2002) Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS* **110**: 263–272.
- Jonasson J, Olofsson M & Monstein HJ (2007) Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. 2002. *APMIS* **115**: 668–677.
- Jureen P, Engstrand L, Eriksson S, Alderborn A, Krabbe M & Hoffner SE (2006) Rapid detection of rifampin resistance in *Mycobacterium tuberculosis* by pyrosequencing technology. *J Clin Microbiol* **44**: 1925–1929.
- Kanduma E, McHugh TD & Gillespie SH (2003) Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *J Appl Microbiol* **94**: 781–791.
- Kang HW, Park DS, Go SJ & Eun MY (2002) Fingerprinting of diverse genomes using PCR with universal rice primers generated from repetitive sequence of Korean weedy rice. *Mol Cells* **13**: 281–287.
- Karjalainen T, Saumier N, Barc MC, Delmee M & Collignon A (2002) *Clostridium difficile* genotyping based on *slpA* variable region in S-layer gene sequence: an alternative to serotyping. *J Clin Microbiol* **40**: 2452–2458.
- Kato H, Yokoyama T & Arakawa Y (2005) Typing by sequencing the *slpA* gene of *Clostridium difficile* strains causing multiple outbreaks in Japan. *J Med Microbiol* **54**: 167–171.
- Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N & Small PM (2001) Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* **11**: 547–554.
- Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ & Hugh-Jones ME (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* **182**: 2928–2936.
- Kellenberger E (2001) Exploring the unknown. The silent revolution of microbiology. *EMBO Rep* **2**: 5–7.
- Keto-Timonen RO, Autio TJ & Korkeala HJ (2003) An improved amplified fragment length polymorphism (AFLP) protocol for

- discrimination of *Listeria* isolates. *Syst Appl Microbiol* **26**: 236–244.
- Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO & Yanofsky C (2000) DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *P Natl Acad Sci USA* **97**: 12170–12175.
- Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**: 275–276.
- Kivioja T, Arvas M, Saloheimo M, Penttilä M & Ukkonen E (2005) Optimization of cDNA-AFLP experiments using genomic sequence data. *Bioinformatics* **21**: 2573–2579.
- Klevytska AM, Price LB, Schupp JM, Worsham PL, Wong J & Keim P (2001) Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* **39**: 3179–3185.
- Koeuth T, Versalovic J & Lupski JR (1995) Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Res* **5**: 408–418.
- Konigsson MH, Bolske G & Johansson KE (2002) Intraspecific variation in the 16S rRNA gene sequences of *Mycoplasma agalactiae* and *Mycoplasma bovis* strains. *Vet Microbiol* **85**: 209–220.
- Korean L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM & Kreiswirth BN (2004) spa typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol* **42**: 792–799.
- Kuhn G, Francioli P & Blanc DS (2007) Double-locus sequence typing using clfB and spa, a fast and simple method for epidemiological typing of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* **45**: 54–62.
- La SB, Elkarkouri K, Li W *et al.* (2008) Rapid comparative genomic analysis for clinical microbiology: the *Francisella tularensis* paradigm. *Genome Res* **18**: 742–750.
- Larson EJ, Hakovirta JR, Cai H, Jett JH, Burde S, Keller RA & Marrone BL (2000) Rapid DNA fingerprinting of pathogens by flow cytometry. *Cytometry* **41**: 203–208.
- Le FP, Fabre M, Denoeud F, Koeck JL & Vergnaud G (2002) High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol* **2**: 37.
- Lee MS, Lee YS & Chiou CS (2006) The suitable restriction enzymes for pulsed-field gel electrophoresis analysis of *Bordetella pertussis*. *Diagn Microb Infect Dis* **56**: 217–219.
- Lefebvre T & Stanhope MJ (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol* **8**: R71.
- Legendre M, Pochet N, Pak T & Verstrepen KJ (2007) Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res* **17**: 1787–1796.
- Li JJ, Pei GL, Pang HX, Bilderbeck A, Chen SS & Tao SH (2006a) A new method for RAPD primers selection based on primer bias in nucleotide sequence data. *J Biotechnol* **126**: 415–423.
- Li W, Chomel BB, Maruyama S, Guptil L, Sander A, Raoult D & Fournier PE (2006b) Multispacer typing to study the genotypic distribution of *Bartonella henselae* populations. *J Clin Microbiol* **44**: 2499–2506.
- Li W, Raoult D & Fournier PE (2007) Genetic diversity of *Bartonella henselae* in human infection detected with multispacer typing. *Emerg Infect Dis* **13**: 1178–1183.
- Li W, Fenollar F, Rolain JM *et al.* (2008) Genotyping reveals a wide heterogeneity of *Tropheryma whippelii*. *Microbiology* **154**: 521–527.
- Liao JC, Li CC & Chiou CS (2006) Use of a multilocus variable-number tandem repeat analysis method for molecular subtyping and phylogenetic analysis of *Neisseria meningitidis* isolates. *BMC Microbiol* **6**: 44.
- Lim EL, Tomita AV, Thilly WG & Polz MF (2001) Combination of competitive quantitative PCR and constant-denaturant capillary electrophoresis for high-resolution detection and enumeration of microbial cells. *Appl Environ Microb* **67**: 3897–3903.
- Lindstedt BA (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* **26**: 2567–2582.
- Lindstedt BA, Heir E, Vardund T & Kapperud G (2000a) Fluorescent amplified-fragment length polymorphism genotyping of *Salmonella enterica* subsp. *enterica* serovars and comparison with pulsed-field gel electrophoresis typing. *J Clin Microbiol* **38**: 1623–1627.
- Lindstedt BA, Heir E, Vardund T, Melby KK & Kapperud G (2000b) Comparative fingerprinting analysis of *Campylobacter jejuni* subsp. *jejuni* strains by amplified-fragment length polymorphism genotyping. *J Clin Microbiol* **38**: 3379–3387.
- Lista F, Faggioni G, Valjevac S *et al.* (2006) Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. *BMC Microbiol* **6**: 33.
- Liu Y & Fratamico P (2006) *Escherichia coli* O antigen typing using DNA microarrays. *Mol Cell Probes* **20**: 239–244.
- Lo YM & Chan KC (2006) Introduction to the polymerase chain reaction. *Method Mol Biol* **336**: 1–10.
- Lockhart DJ, Dong H, Byrne MC *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**: 1675–1680.
- Lucchini S, Thompson A & Hinton JC (2001) Microarrays for microbiologists. *Microbiology* **147**: 1403–1414.
- Lupski JR & Weinstock GM (1992) Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* **174**: 4525–4529.
- Maiden MC, Bygraves JA, Feil E *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *P Natl Acad Sci USA* **95**: 3140–3145.

- Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Matsheka MI, Lastovica AJ, Zappe H & Elisha BG (2006) The use of (GTG)₅ oligonucleotide as an RAPD primer to type *Campylobacter concisus*. *Lett Appl Microbiol* **42**: 600–605.
- Medini D, Donati C, Tettelin H, Massignani V & Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589–594.
- Melles DC, van Leeuwen WB, Snijders SV, Horst-Kreft D, Peeters JK, Verbrugh HA & van BA (2007) Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. *J Microbiol Meth* **69**: 371–375.
- Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM & Mahairas GG (2004) The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *J Bacteriol* **186**: 7123–7133.
- Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE & Ecker JR (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1–15.
- Mokrousov I, Vyazovaya A, Kolodkina V, Limeschenko E, Titov L & Narvskaya O (2008) Novel macroarray-based method of *Corynebacterium diphtheriae* genotyping: evaluation in a field study in Belarus. *Eur J Clin Microbiol*, Dec 17. [Epub ahead of print] PMID: 19089478.
- Mollet C, Drancourt M & Raoult D (1997) rpoB sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26**: 1005–1011.
- Monot M, Honore N, Baliere C, Ji B, Sow S, Brennan PJ & Cole ST (2008) Are variable-number tandem repeats appropriate for genotyping *Mycobacterium leprae*? *J Clin Microbiol* **46**: 2291–2297.
- Mortimer P & Arnold C (2001) FAFLP: last word in microbial genotyping? *J Med Microbiol* **50**: 393–395.
- Mostrom P, Gordon M, Sola C, Ridell M & Rastogi N (2002) Methods used in the molecular epidemiology of tuberculosis. *Clin Microbiol Infect* **8**: 694–704.
- Mueller UG & Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends Ecol Evol* **14**: 389–394.
- Muyzer G, de Waal EC & Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microb* **59**: 695–700.
- Northey G, Gal M, Rahmati A & Brazier JS (2005) Subtyping of *Clostridium difficile* PCR ribotype 001 by REP-PCR and PFGE. *J Med Microbiol* **54**: 543–547.
- Nyren P (1987) Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal Biochem* **167**: 235–238.
- Ogata H, Audic S, Barbe V, Artiguenave F, Fournier PE, Raoult D & Claverie JM (2000) Selfish DNA in protein-coding genes of *Rickettsia*. *Science* **290**: 347–350.
- Otal I, Martin C, Vincent-Levy-Frebault V, Thierry D & Gicquel B (1991) Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. *J Clin Microbiol* **29**: 1252–1254.
- Poulet S & Cole ST (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* **163**: 87–95.
- Power EG (1996) RAPD typing in microbiology – a technical review. *J Hosp Infect* **34**: 247–265.
- Price EP, Smith H, Huygens F & Giffard PM (2007) High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. *Appl Environ Microb* **73**: 3431–3436.
- Ransom G & Kaplan B (1998) USDA uses PulseNet for food safety. *J Am Vet Med Assoc* **213**: 1107.
- Rappe MS & Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369–394.
- Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B & Barrett TJ (2006) Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog Dis* **3**: 59–67.
- Rombauts S, Van De PY & Rouze P (2003) AFLP in Silico, simulating AFLP fingerprints. *Bioinformatics* **19**: 776–777.
- Ronaghi M, Uhlen M & Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* **281**: 363–365.
- Roth SB, Jalava J, Ruuskanen O, Ruohola A & Nikkari S (2004) Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections. *J Clin Microbiol* **42**: 4268–4274.
- Ruiz-Garbajosa P, Bonten MJ, Robinson DA *et al.* (2006) Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *J Clin Microbiol* **44**: 2220–2228.
- Sacchi CT, Whitney AM, Reeves MW, Mayer LW & Popovic T (2002) Sequence diversity of *Neisseria meningitidis* 16S rRNA genes and use of 16S rRNA gene sequencing as a molecular subtyping tool. *J Clin Microbiol* **40**: 4520–4527.
- Sadeghifard N, Gurtler V, Beer M & Seviour RJ (2006) The mosaic nature of intergenic 16S–23S rRNA spacer regions suggests rRNA operon copy number variation in *Clostridium difficile* strains. *Appl Environ Microb* **72**: 7311–7323.
- Salser WA (1974) DNA sequencing techniques. *Annu Rev Biochem* **43**: 923–965.
- Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci USA* **74**: 5463–5467.
- Schena M, Shalon D, Davis RW & Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schork NJ, Fallin D & Lanchbury JS (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* **58**: 250–264.
- Schwartz DC & Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**: 67–75.

- Sciacchitano CJ (1998) DNA fingerprinting of *Listeria monocytogenes* using enterobacterial repetitive intergenic consensus (ERIC) motifs-polymerase chain reaction/capillary electrophoresis. *Electrophoresis* **19**: 66–70.
- Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, Bost DA, Riehman M, Naidich S & Kreiswirth BN (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* **37**: 3556–3563.
- Shutt CK, Pounder JI, Page SR, Schaecher BJ & Woods GL (2005) Clinical evaluation of the DiversiLab microbial typing system using repetitive-sequence-based PCR for characterization of *Staphylococcus aureus* strains. *J Clin Microbiol* **43**: 1187–1192.
- Snelling AM, Gerner-Smith P, Hawkey PM, Heritage J, Parnell P, Porter C, Bodenham AR & Inglis T (1996) Validation of use of whole-cell repetitive extragenic palindromic sequence-based PCR (REP-PCR) for typing strains belonging to the *Acinetobacter calcoaceticus*–*Acinetobacter baumannii* complex and application of the method to the investigation of a hospital outbreak. *J Clin Microbiol* **34**: 1193–1202.
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503–517.
- Spratt BG (1999) Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr Opin Microbiol* **2**: 312–316.
- Stackebrandt E & Goebel BM (1994) Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846–849.
- Stern MJ, Ames GE, Smith NH, Robinson EC & Higgins CF (1984) Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37**: 1015–1026.
- Swaminathan B, Gerner-Smith P, Ng LK, Lukinmaa S, Kam KM, Rolando S, Gutierrez EP & Binsztein N (2006) Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. *Foodborne Pathog Dis* **3**: 36–50.
- Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH & 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.
- Tenover FC, Vaughn RR, McDougal LK, Fosheim GE & McGowan JE Jr (2007) Multiple locus variable number tandem repeat assay analysis of methicillin-resistant *Staphylococcus aureus* strains. *J Clin Microbiol* **45**: 2215–2219.
- Tettelin H, Massignani V, Cieslewicz MJ *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *P Natl Acad Sci USA* **102**: 13950–13955.
- Thibodeau SN (1987) Use of restriction fragment length polymorphism analysis for detecting carriers of “fragile X” syndrome. *Clin Chem* **33**: 1726–1730.
- Thompson JR, Randa MA, Marcelino LA, Tomita-Mitchell A, Lim E & Polz MF (2004) Diversity and dynamics of a north atlantic coastal *Vibrio* community. *Appl Environ Microb* **70**: 4103–4110.
- Todd R, Donoff RB, Kim Y & Wong DT (2001) From the chromosome to DNA: restriction fragment length polymorphism analysis and its clinical application. *J Oral Maxil Surg* **59**: 660–667.
- Trad S, Allignet J, Frangeul L *et al.* (2004) DNA microarray for identification and typing of *Staphylococcus aureus* isolates. *J Clin Microbiol* **42**: 2054–2064.
- Tyler KD, Wang G, Tyler SD & Johnson WM (1997) Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens. *J Clin Microbiol* **35**: 339–346.
- Unemo M, Olcen P, Jonasson J & Fredlund H (2004) Molecular typing of *Neisseria gonorrhoeae* isolates by pyrosequencing of highly polymorphic segments of the *porB* gene. *J Clin Microbiol* **42**: 2926–2934.
- U'Ren JM, Schupp JM, Pearson T *et al.* (2007) Tandem repeat regions within the *Burkholderia pseudomallei* genome and their application for high resolution genotyping. *BMC Microbiol* **7**: 23.
- van BA (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunol Med Mic* **49**: 22–27.
- van BA, Scherer S, van AL & Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol R* **62**: 275–293.
- van BA, Struelens M, de VA, Verbrugh H & Tibayrenc M (2001) Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin Microbiol Rev* **14**: 547–560.
- van den Berg RJ, Schaap I, Templeton KE, Klaassen CH & Kuijper EJ (2007) Typing and subtyping of *Clostridium difficile* isolates by using multiple-locus variable-number tandem-repeat analysis. *J Clin Microbiol* **45**: 1024–1028.
- van Embden JD, van GT, Kremer K, Jansen R, van Der Zeijst BA & Schouls LM (2000) Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol* **182**: 2393–2401.
- Vergnaud G & Denoeud F (2000) Minisatellites: mutability and genome architecture. *Genome Res* **10**: 899–907.
- Versalovic J, Koeuth T & Lupski JR (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**: 6823–6831.
- Versalovic J, Kapur V, Mason EO Jr, Shah U, Koeuth T, Lupski JR & Musser JM (1993) Penicillin-resistant *Streptococcus pneumoniae* strains recovered in Houston: identification and molecular characterization of multiple clones. *J Infect Dis* **167**: 850–856.
- Vogel L, Jones G, Triep S, Koek A & Dijkshoorn L (1999) RAPD typing of *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Serratia marcescens* and *Pseudomonas aeruginosa* isolates using standardized reagents. *Clin Microbiol Infect* **5**: 270–276.

- Vos P, Hogers R, Bleeker M, Reijans M, van de LT, Hornes M, Frijters A, Pot J, Peleman J & Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**: 4407–4414.
- Wahab T, Hjalmarsson S, Wollin R & Engstrand L (2005) Pyrosequencing *Bacillus anthracis*. *Emerg Infect Dis* **11**: 1527–1531.
- Wayne LG, Brenner DJ, Colwell RR *et al.* (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**: 463–464.
- Welsh J & McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* **18**: 7213–7218.
- Werner G, Klare I & Witte W (2007) The current MLVA typing scheme for *Enterococcus faecium* is less discriminatory than MLST and PFGE for epidemic-virulent, hospital-adapted clonal types. *BMC Microbiol* **7**: 28.
- Wichelhaus TA, Hunfeld KP, Boddingtonhaus B, Kraiczy P, Schafer V & Brade V (2001) Rapid molecular typing of methicillin-resistant *Staphylococcus aureus* by PCR-RFLP. *Infect Cont Hosp Ep* **22**: 294–298.
- Williams JG, Kubelik AR, Livak KJ, Rafalski JA & Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* **18**: 6531–6535.
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Woods CR, Versalovic J, Koeuth T & Lupski JR (1993) Whole-cell repetitive element sequence-based polymerase chain reaction allows rapid assessment of clonal relationships of bacterial isolates. *J Clin Microbiol* **31**: 1927–1931.
- Zhan Y & Kulp D (2005) Model-P: a base calling method for resequencing microarrays of diploid samples. *Bioinformatics* **21**: ii182–ii189.
- Zhang SL, Shen JG, Xu PH, Li DX, Sun ZQ, Li L, Yang ZR & Sun Q (2007) A novel genotypic test for rapid detection of multidrug-resistant *Mycobacterium tuberculosis* isolates by a multiplex probe array. *J Appl Microbiol* **103**: 1262–1271.
- Zhao S, Mitchell SE, Meng J, Kresovich S, Doyle MP, Dean RE, Casa AM & Weller JW (2000) Genomic typing of *Escherichia coli* O157:H7 by semi-automated fluorescent AFLP analysis. *Microbes Infect* **2**: 107–113.
- Zhou J (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288–294.
- Zhu Y, Fournier PE, Ogata H & Raoult D (2005) Multispacer typing of *Rickettsia prowazekii* enabling epidemiological studies of epidemic typhus. *J Clin Microbiol* **43**: 4708–4712.
- Zwick ME, McAfee F, Cutler DJ, Read TD, Ravel J, Bowman GR, Galloway DR & Mateczun A (2005) Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol* **6**: R10.