

A Phylogenetic Perspective on Molecular Epidemiology

Xin Wang¹, I. King Jordan² and Leonard W. Mayer¹

¹Centers for Disease Control and Prevention, Atlanta, GA, USA, ²Georgia Institute of Technology, Atlanta, GA, USA

INTRODUCTION

Epidemiology is the study of disease prevalence and distribution in population and risk factors that are associated with a disease. This discipline plays important roles in public health surveillance and disease control, and has improved our understanding of host–pathogen interaction and disease transmission. In addition to identifying the aetiology of a disease, epidemiology also assesses risk factors such as environmental and host genetic factors and how they contribute to the burden of the disease. Association between risk factors and infectious disease at genetic level is defined by molecular epidemiology [1,2]. How to define molecular epidemiology has been a controversial issue since the 1970s [3]. It is generally accepted that molecular epidemiology is a discipline that resolves epidemiological problems using molecular techniques [1–3]. Molecular epidemiology employs molecular or genetic markers to trace the development of a disease in a population and to understand transmission, as well as the population structure and evolution of bacterial pathogens [4]. This discipline has grown rapidly in the past couple of decades with the advances in DNA-based molecular typing techniques, which are currently available for a number of pathogens such as bacteria, viruses and fungi. More recently this has also included the complete genomes of these pathogens. Molecular typing measures genetic difference between strains in a population; phylogenetic analysis classifies them into different genetic groups or clusters based on genetic relationship and diversity [5–7]. Phylogenetic analysis of molecular typing data allows the determination of the genetic relatedness of strains from different sources, geographic locations and/or even different time periods and inferring evolutionary relationships. If bacterial strains from different patients

are closely related to each other at genetic level, it suggests these strains have a common origin and epidemiological linkage between these cases. Information obtained from molecular typing and phylogenetic analysis is valuable for outbreak and epidemic investigations. Bacterial strains causing outbreaks can spread rapidly and undergo genetic variations during the spread. Phylogenetic analysis of genetic profiles of these strains determines whether they are from the same clone or if new clones have emerged. This is extremely important for pathogens with decreased antibiotic susceptibility or causing vaccine-preventable disease when strain replacement occurs. Monitoring of genetic variations that affect antibiotic susceptibility or vaccine targets helps to determine appropriate public health interventions for control and prevention of infectious disease. Molecular epidemiology also provides an effective approach to monitor and track strains with increased virulence or transmissibility, and to study bacterial population biology and evolution [3].

It should be emphasized that, as the cornerstone of molecular epidemiology, an effective molecular typing scheme should provide sufficient discriminatory power, be reproducible among different laboratories, and be easily performed and standardized. Ideally the data should be able to be digitized to be stored in a publicly accessible database and be easily transmitted between laboratories. A number of molecular typing schemes including one or more gene targets have been evaluated for differentiating bacterial strains. Each typing scheme is designed to measure different types of genotypic or phenotypic variations of a bacterial species. Multiple typing schemes may be available for a particular bacterial species. It is important to keep in mind that selection of molecular typing scheme depends on pathogens to be investigated and questions to be addressed [6]. Short-term or local epidemiology

focuses on genetic variations or microvariations that may occur during an individual outbreak or epidemic event in a particular geographic location. The goal of long-term or global epidemiology is to measure the accumulation of genetic variation overtime and establish a linkage between strain lineage and disease on a global scale [8].

MOLECULAR TYPING METHODS INDEXING GENETIC VARIATIONS

Early typing techniques mostly rely on the DNA banding patterns generated by restriction enzyme digestion to determine genetic relatedness of bacterial strains. Plasmid analysis is probably the first molecular typing tool to differentiate bacterial strains by examining strains for presence/absence, size, and restriction digestion patterns of plasmid DNA [9,10]. Other restriction enzyme digestion-based techniques include restriction fragment length polymorphism analysis (RFLP), ribotyping and pulsed-field gel electrophoresis (PFGE) [10,11]. While RFLP and ribotyping have limited applications in recent epidemiology investigations, PFGE remains one of the most common typing tools for a variety of bacterial species [12]. In the genomic era, molecular typing technology has evolved to newer-generation typing tools and is increasingly relying on PCR amplification- or DNA sequencing-based methods. Before DNA sequencing technology became popular, PCR-based typing schemes such as amplified fragment length polymorphisms (AFLP), random amplified polymorphic DNA PCR (RAPD-PCR), repetitive element PCR (Rep-PCR) and MLVA were commonly used in microbiology laboratories. These methods are relatively fast, inexpensive and easy to perform. With rapid expansion of next-generation DNA sequencing technology (e.g. 454, Illumina and PacBio) in the past decade, sequencing-based typing methods have become more and more popular because they are able to unambiguously determine each mutation within a particular target gene. The sequencing-based typing data are reproducible, portable and allow comparison of the results among different laboratories worldwide, which are essential to study global epidemiology of infectious disease [5,7,10]. With the steady decrease in the cost for DNA sequencing and increased access to bioinformatics tools, sequencing-based typing is expected to soon be widely implemented in research and clinical microbiology laboratories.

DNA-based molecular typing methods commonly involve one or more target genes or the uncharacterized genome. Research has demonstrated that typing methods including multiple target genes improve strain typing at species and subspecies levels [13–16]. This chapter reviews several common typing schemes currently in use

for bacterial strain typing (Table 29.1), including the basic principles of these techniques, management and analysis of typing data and a few example applications.

DNA SEQUENCING-BASED METHODS

16S rRNA Gene Sequencing

The 16S ribosomal RNA gene codes for the RNA component of the 30S subunit of the bacterial ribosome. It is widely present in all bacterial species. Different bacterial species have one to multiple copies of the 16S rRNA gene. 16S rRNA gene sequencing is by far one of the most common methods targeting housekeeping genes to study bacterial phylogeny and genus/species classification [17]. DNA–DNA hybridization is the gold standard for identifying bacterial species [18]. Because of the complexity of DNA–DNA hybridization, 16S rRNA gene sequencing is used as a tool to identify bacteria at the species level and assist with differentiating between closely related bacterial species [8]. Many clinical laboratories rely on this method to identify unknown pathogenic strains [19]. During 2001–2007, 16S rRNA gene sequencing identified 215 novel bacterial species, 29 of which were classified to novel genera [20]. According to proposed guidelines for bacterial classification, strains with less than 97% similarity in 16S rRNA gene sequence represent different bacterial species; the ones that show more than 97% similarity should be classified using an alternative approach [21].

While this techniques is highly useful for genus classification and able to resolve more than 90% of the isolates, it is sometimes less valuable for species classification because of low discriminatory power, with only 65–83% of strains being identified and the rest remaining unresolved [21]. In some cases, strains showing more than 99% similarity in 16S rRNA gene sequences are actually distinguishable at species level by DNA–DNA hybridization. A good example comes from the type strains of *B. globisporus* and *B. psychrophilus* that show 99.5% sequence similarity in 16S rRNA genes, but only 23–50% genetically associated by DNA–DNA hybridization [22]. In addition to being used as a tool for identification, 16S rRNA gene relatedness has been used as a subtyping method. Sacchi et al. and others demonstrated the utility of 16S typing for *Neisseria meningitidis*, *Haemophilus influenzae* and *Bacillus anthracis* [23–26]. Using 16S rRNA gene sequencing, Anderson et al. identified seven invasive cases caused by *Haemophilus haemolyticus*, which were previously misidentified as *H. influenzae*. They also noted that despite the fact that all seven *H. haemolyticus* strains form the same phylogenetic

TABLE 29.1 Molecular Typing Methods

Typing Method	Basic Features	Public Database Link
16S rRNA gene sequencing	DNA sequencing based; common method for bacteria phylogeny and genus/species classifications	http://rdp.cme.msu.edu/
Multilocus sequence typing	DNA sequencing based; standard strain typing method including six to eight loci; typing scheme has been developed for many bacterial species	http://pubmlst.org http://www.mlst.net
Ribosomal multilocus sequence typing (rMLST)	DNA sequencing based; strain typing tool recently developed based on 53 loci coding for ribosomal proteins; rMLST profiles can be obtained from BIGSdb with a short turnaround	http://pubmlst.org/software/database/bigsgdb/
Genome-based typing	DNA sequencing-based tool for typing/classification using bacterial core genes; detects genetic variations at the genome level and provides higher resolution than other common typing methods	http://www.ncbi.nlm.nih.gov/genome/ http://img.jgi.doe.gov
Single nucleotide polymorphism	DNA sequencing-based typing tool targeting polymorphic sequences of the whole or partial genome sequences; valuable for tracking the spread of monomorphic pathogens	
Multiple-locus variable-number tandem repeat analysis	Amplification-based method; determines the number of tandem repeats in multiple loci; capable of detecting genetic differences between strains of highly homogeneous species	http://www.pasteur.fr/mlva http://minisatellites.u-psud.fr/MLVAnet http://www.pulsenetinternational.org/protocols/Pages/mlva.aspx
Amplified fragment length polymorphisms	Amplification-based method; analyses a subset of DNA regions of restriction enzyme-digested bacterial genome	NA
Random amplified polymorphic DNA PCR or arbitrarily primed PCR	Amplification-based method; target DNA regions randomly distributed in bacterial genome	NA
Repetitive extragenic palindromic PCR	Amplification-based method; detects DNA regions between two repetitive DNA sequences in close proximity and creates DNA banding profiles; a semiautomated REP-PCR system (DiversiLab) is available and allow interlaboratory comparison; used for many bacterial species	NA
Pulsed-field gel electrophoresis	Restriction digestion-based method; commonly used for epidemiological investigations of foodborne diseases and others	www.pulsenetinternational.org

cluster with other known *H. haemolyticus*, one known *H. haemolyticus* strain was outside of this cluster, indicating the ambiguity of using 16S rRNA for strain classification [27,28]. For the same reason, this method is not optimal for typing strains of the same species [15].

Multilocus Sequence Typing (MLST)

MLST is based on principles similar to multilocus enzyme electrophoresis (MLEE), where genetic variations of bacterial strains are inferred from comparison of the electrophoretic mobility of housekeeping enzymes. Strains are classified into different electrophoretic types, which are further grouped into clones and subgroups. As there are only limited variations with a locus, the resolution of MLEE is low; high resolution can only be achieved by analysing 20 or more loci, which is time-consuming and labour-intensive. Compared to MLEE, MLST analysis requires fewer loci, targeting the internal fragment of only seven housekeeping genes with a length of 400–500 bp. Each DNA fragment is sequenced by a pair of forward and reverse primers to ensure accurate determination of mutations in the seven loci. For each locus, an allele number is assigned to each unique nucleotide sequence. A sequence type is defined by the allelic profile of the seven loci. Strains that have at least four alleles in common are usually defined as the same clonal complex. DNA sequencing data are transferrable and reproducible among different laboratories. The method has high throughput and is scalable for population-based studies.

MLST was first developed and implemented for defining global epidemiology of meningococcal strains [8,29,30]. The typing scheme initially included six loci and was subsequently modified to use seven loci to improve discrimination power [30]. The seven genes included in this typing scheme are *abcZ* (putative ABC transporter), *adk* (adenylate kinase), *aroE* (shikimate dehydrogenase), *fumC* (fumarase), *gdh* (glucose-6-phosphate dehydrogenase), *pdhC* (pyruvate dehydrogenase subunit) and *pgm* (phosphoglucosmutase). Although horizontal gene transfer occurs frequently among *N. meningitidis* strains, genetic mutations accumulate slowly in housekeeping genes to maintain their biological functions. Therefore, the seven loci are considerably stable over time which allows researchers to study global epidemiology using this typing scheme. To date, MLST typing schemes based on six to eight loci have been developed for 95 bacterial species (<http://pubmlst.org>) and are considered as standard typing methods for many species. However, MLST may be less discriminative than other typing methods for local outbreak investigations. For example, *N. meningitidis* strains from different outbreak episodes that belonged to the same clonal complex or even

the same sequence type by MLST analysis may form distinct phylogenetic clusters using alternative typing methods [31]. This further highlights the importance of selecting proper typing tools for epidemiological investigation [6].

A web-based public MLST database has been created to store all allelic profiles for many different bacterial species. New alleles will be assigned through submission to this web-based database. MLST data of all bacterial species are currently available at <http://pubmlst.org> that is hosted at the University of Oxford, UK; Imperial College, London, UK; the Environmental Research Institute, Cork, Ireland; and the Pasteur Institute, Paris, France and <http://www.mlst.net>. This database includes thousands of allelic profiles from numerous isolates of different bacterial species from diverse geographical locations. The data from both invasive and carriage isolates are captured in the MLST database for easy comparison at the genetic level. So far, the *N. meningitidis*, *Streptococcus pneumoniae* and *Campylobacter jejuni* databases are currently the three largest MLST databases.

Ribosomal Multilocus Sequence Typing (rMLST)

Increased evidence has suggested that typing methods typically relying on a single or a few markers do not provide sufficient resolution power and, therefore, may not be reliable for typing strains among closely related bacteria as these genes only account for a tiny portion of the bacterial genome [13,32]. In recent years, efforts have been made to develop genome-based typing methods embracing the full genome sequence (see ‘Genome-Based Typing’ section below). With the increase in the number of bacterial genome sequences, comparative genomic analysis becomes possible for interrogating the genomic data and identifying new typing markers. Jolly et al. recently developed a ribosomal multilocus sequence typing method that indexes genetic variations of ribosomal protein-encoding genes [15]. Ribosomal genes have historically been used for studying molecular epidemiology and phylogenetic relatedness of bacterial strains. Examples include 16S rRNA gene sequencing and ribotyping which target one or more ribosomal RNA genes [33]. Although rRNA genes may not be perfect markers to infer bacterial evolution, they are sound predictors for indexing bacterial phylogenetic relationships [34]. The newly developed rMLST is expected to provide higher resolution power than other ribosomal gene-based methods.

As ribosomal protein-encoding genes are present in all bacterial species, rMLST can be potentially used for classifying and typing all bacteria. Jolly et al. took advantage of genome data from 1902 isolates in the Integrated

Microbial Genomes (IMG) database (<http://img.jgi.doe.gov/>) and identified 53 loci coding for ribosomal proteins for discriminating bacterial strains from *Bacillus*, *Listeria*, *Streptococcus*, etc. [15]. rMLST is considered as an extension of the 16S rRNA gene sequencing method. Although some loci may be more discriminatory for a particular bacterial species than others, these 53 loci provide a framework for developing rMLST typing schemes for other bacterial species. Selection of rMLST loci for a typing scheme depends on whether the scheme will be used for species- or subspecies/strain-typing. rMLST typing of 53 loci is proven to discriminate *Streptococcus* strains at species and subspecies levels. However, for highly homogeneous species, also defined as genetically monomorphic species, such as *B. anthracis* and *Yersinia pestis*, sequencing of these ribosomal protein-encoding genes may not be sufficiently discriminatory, and therefore fine-typing methods offering more detailed analysis of bacterial genetic variations may be necessary to achieve expected resolution power.

Bacterial Isolate Genome Sequence Database (BIGSdb) provides a useful platform for storage and analysis of bacterial DNA sequence data ranging from a single gene to genome sequences. To date, the database contains genome sequences from a total of 19 763 bacterial isolates collected from a broad range of sources. It offers rapid analysis and is able to assemble genome and obtain rMLST profiles in about 1–2 hours [15,35].

Genome-Based Typing

Science has entered into a new era of genomics. Genomic analysis has revealed tremendous diversity of bacterial genomes and provided invaluable information for studying bacterial evolution, defining population structure, discovering new therapeutic target and vaccine candidates [36–39]. Bacterial genomes are relatively small, with a size of several mega base pairs. Many bacterial species are naturally competent and able to acquire exogenous DNA through intra- or interspecies horizontal gene transfer (HGT). Therefore, bacterial genomes are relatively diverse and contain a high proportion of dispensable sequences, possibly introduced through frequent HGT [36,40]. Pioneer studies have proposed a new concept ‘pan-genome’ which includes ‘core genes’ that are present in all members of a bacterial species and ‘distributed or dispensable genes’, which are unique to one or more members of a bacterial species but not shared by all of the members. Core genes encode basic biological functions such as metabolism, chromosomal replication, transcription and translation, cell wall synthesis and other common phenotypic traits. The dispensable sequences confer specific functionality and have been suggested to link to virulence, functions for adaption to environment and other

strain-specific traits [14,39,41]. Due to the complexity of the bacterial gene repertoire, it is not surprising that the number of core genes or dispensable genes varies with increased genomes sequenced. Tettelin et al. used a mathematic model to estimate the number of core genes that are present in Group B streptococci (GBS) strains. The extrapolation curve indicates that core genes of GBS strains can reach 1806 and remain constant [39]. The number of dispensable genes will continue increasing with the increase in the number of sequenced genomes. While the dispensable genes are beneficial for bacterial adaption and fitness, they prevent the establishment of phylogenetic relationship between bacterial species and between strains within a particular bacterial species. Therefore, genome-based typing should focus on the core genes that are universally present among strains of a particular bacterial species to address questions pertaining to bacterial phylogeny. Inclusion of dispensable sequences in typing schemes may alter the structure of phylogenetic trees and lead to incorrect interpretation of phylogeny [34,36,42].

Genome sequencing technology in the early days was based on Sanger sequencing, which is very time-consuming and costly. *H. influenzae* Rd genome was the first complete bacterial genome that was sequenced [43]. The project took about a year and cost approximately one million dollars to complete [16]. Emergence of new next-generation sequencing platforms such as Roche 454 GS FLX, Illumina Genome Analyzer and PacBio RS remarkably decreased the cost and improved the turnaround time of genome sequencing. With the new platforms, a bacterial genome can be sequenced in hours, with the cost as low as \$25 per mega base pair [44]. It is now increasingly feasible for clinical and public health laboratories to study bacterial genomes at the population level using the low-cost, high-throughput whole-genome sequencing techniques. A recent study reported a hospital microbiology laboratory that did genomic sequencing on all 130 samples from 116 patient specimens representing the yield of an entire day [45].

One of the major challenges using this technology in microbiology laboratories is handling the enormous amount of data generated from these platforms. Assembly and annotation of genomic sequences require broad knowledge in bioinformatics and the process can be extremely slow. To accelerate this process, a number of tools have been developed for genome assembly, gene prediction and annotation, and data storage. Web-based tools have also been developed for comparative genomic analysis to simplify the procedures. Many of these tools, such as Integrated Microbial Genomes, are publicly accessible. Each tool has its strength and weakness; caution should be taken when researchers select proper tools [15,46,47].

As all genetic information is contained with genome sequences, genome-based typing is able to detect all genetic variations around the genome at nucleotide level and provides higher resolution than some of the common typing methods [14,16]. This is extremely important for monomorphic organisms such as *Y. pestis*, *S. typhi* and *Mycobacterium tuberculosis*. Because strains of monomorphic species have a low degree of genetic diversity, typing tools relying on a few housekeeping genes or selected target genes do not provide sufficient resolution for strain differentiation. Efforts are being made to develop typing schemes targeting core genomic genes or genome-wide single nucleotide polymorphism loci (see section on single nucleotide polymorphism typing below). Lerat et al. developed a genome-based approach to resolve the phylogeny of the γ -Proteobacteria exhibiting high frequency of horizontal gene transfer. The study demonstrates that gene orthologues among γ -Proteobacteria are able to yield congruent information for inferring organismal phylogeny. Genome-based typing has been used to investigate outbreaks and epidemics caused by a number of pathogens [4]. Comparative genomics analysis between *V. cholerae* strains that caused the 2010 cholera outbreak in Haiti and strains from different geographic sources confirmed the Haitian *V. cholerae* strains are distinct from the strains from Latin America and the US Gulf Coast, but genetically close to strains from South Asia, with highest similarity to strains from Nepal [48,49]. These studies provided strong evidence that the Haitian cholera outbreak strains were introduced to Haiti from Nepal through human activities as suggested by the previous epidemiological investigation [50].

SINGLE NUCLEOTIDE POLYMORPHISM (SNP) TYPING

Single nucleotide polymorphism (SNP) probably represents the most abundant genetic variations in the genome. SNP typing was initially used for detection of polymorphic DNA sequences in eukaryotic genomes in order to establish linkage between gene loci and certain disease conditions and has recently been adopted for bacterial strain typing. Genome-wide SNPs are identified by searching throughout the bacterial genome and validated for discriminating strains. Some SNP-based typing methods target specific genes or DNA regions. In order to provide high discriminatory power, the loci selected for SNP typing should have a considerably higher level of polymorphism. As SNP typing is a relatively new technique for bacterial genotyping, the results should be compared with other typing methods to determine whether the selected loci are suitable for SNP typing. Methods to detect polymorphism within these loci include DNA sequencing or pyrosequencing, mass spectrometry and

real-time PCR. Sequencing-based techniques provide accurate information with regards to the nucleotide polymorphism at defined locations but are more expensive compared to other methods [51].

SNP-based typing has proven to be highly valuable for tracking the spread of monomorphic pathogens during outbreaks and epidemics and for reconstruction of evolutionary history [52,53]. A phylogenetic tree constructed on a set of 933 SNPs identified from 286 *Y. pestis* isolates revealed several geographic-specific lineages represented by phylogenetic branches in the tree [54]. It was inferred that *Y. pestis* evolved in China more than 2600 years ago as isolates from China scattered over all four phylogenetic branches, and was then transmitted to other geographic locations such as Central/South Africa, Madagascar and the USA through multiple occasions. The time that these transmissions occurred can also be inferred from the SNP-based tree. Although other typing methods such as PFGE, MLVA and CRISPR (clustered regulatory short palindromic repeats) were employed to discriminate strains of monomorphic species, studies have suggested that those methods sometimes led to incorrect interpretation of phylogenetic relationships among strains examined and SNP typing based on robust markers is more reliable for defining discrete strain lineages and for epidemiological investigations [13,52,55].

AMPLIFICATION-BASED METHODS

Multiple-Locus Variable-Number Tandem Repeat Analysis (MLVA)

MLVA is a relatively new molecular typing technique employed to subtype bacterial strains for epidemiological investigations. This method detects the copy numbers of repeated DNA sequences that are dispersed throughout the bacterial genome. As the number of these tandem repeats in a particular locus varies from strain to strain, they are known as variable-number tandem repeats or VNTRs [56]. For a particular bacterium, VNTRs may be present in multiple [4–15] loci or regions. Specific loci that are unique to a particular bacterial species are selected as MLVA markers. In brief, the VNTR loci are first PCR amplified; PCR products are subsequently separated on an agarose gel or an automated capillary DNA sequencer. The number of tandem repeats is assessed based on the size of the PCR products. Similar to MLST, the MLVA profile is defined by the number of repeats of the VNTR loci. Each unique MLVA profile coded by a multidigit is assigned a MLVA type number (e.g. MT21). The profile codes can be stored into a database for strain comparison and epidemiological studies. Phylogenetic clusters of the related MLVA types are often classified into MLVA complexes. This typing method is able to

detect genetic differences between strains of highly homogeneous species such as *Mycobacterium tuberculosis* [57].

The first study on the use of MLVA was reported in 1997 by Van Belkum et al. [58]. The study examined eight VNTR regions with repeats of 3–6 nucleotides, and determined that four regions are suitable for subtyping *H. influenzae* strains because of their genetic stability. Using this typing scheme, *H. influenzae* serotype b strains collected from an outbreak were proven to be clonal. Since then, MLVA has been increasingly used to study a variety of bacterial pathogens including *Staphylococcus aureus*, *Y. pestis*, *B. anthracis*, *Salmonella enterica*, and the list continues to grow [57–61].

Several web-based MLVA databases and analysis platforms are currently in use, including MLVA-Net hosted by the Pasteur Institute, France (www.pasteur.fr/mlva), MLVA bank hosted by the University of Orsay, France (<http://minisatellites.u-psud.fr/MLVANet/>), and PulseNet as part of PulseNet International network (<http://www.pulsenetinternational.org/protocols/Pages/mlva.aspx>). The minisatellite database was the first VNTR database specifically created for bacterial species [62]. These platforms allow comparing MLVA profiles of strains worldwide and determining geographic and temporal distribution of MLVA types of bacterial pathogens. Development of the MLVA typing scheme has been made achievable because of the advances in genome sequencing and the computing platforms available to assist with screening bacterial genomes for potential VNTRs [57].

Amplified Fragment Length Polymorphisms (AFLP)

AFLP identifies genetic differences between two bacterial genomes using a combination of restriction enzyme digestion and powerful PCR. In this method, DNA chromosomal genomes are first digested with one or two restriction endonucleases. When two enzymes are being used, one often cleaves a bacterial genome more frequently than the other. The resulting genomic fragments are then ligated to adapters with known sequences and amplified using PCR primers that are designed to anneal to the sequence of the adapters, the remaining restriction site sequence and one or more nucleotides from the genomic fragments generated from the restriction digestion. Inclusion of one or more nucleotides from genomic fragments reduces the number of genomic fragments being amplified and analysed. This feature allows AFLP to analyse a subset of DNA regions randomly scattered throughout the genome. PCR products are resolved using agarose electrophoresis and visualized under UV light. A different version of AFLP (fluorescent AFLP) was recently developed, which uses primers labelled with fluorescent dyes;

labelled PCR products can be resolved by capillary electrophoresis [63]. The DNA fingerprint patterns of bacterial strains are compared to determine their genetic or phylogenetic relationship [64]. In order to achieve acceptable resolution, different primer combinations need to be tested to define the subset of DNA regions that serve as AFLP makers for a particular bacterial species.

Random Amplified Polymorphic DNA PCR (RAPD-PCR) or Arbitrarily Primed PCR (AP-PCR)

Unlike other PCR-based methods, which rely on restriction digestion of the bacterial genome to generate amplification profiles, RAPD-PCR and AP-PCR amplify DNA regions randomly distributed in the bacterial genome using short arbitrary primers (normally 6–10 nucleotides for RAPD-PCR and 20–34 nucleotides for AP-PCR) that do not target any specific genomic sequences. Under low-stringency conditions, these arbitrary primers tolerate mismatches and allow amplification of random genomic regions. AP-PCR requires high annealing temperature for optimal amplification following the low annealing temperature in the first few cycles. PCR products are resolved on agarose gel electrophoresis or capillary electrophoresis and generate a unique DNA fingerprint. The fingerprint patterns can be compared between different bacterial strains to infer their genetic relationship. Both PCR techniques are easy to perform and do not require any knowledge of bacterial genetic compositions. However, one of the major drawbacks of these methods is that even slight changes in cycling conditions or reagents will generate variations in DNA banding patterns, which makes it difficult to compare the banding pattern of bacterial strains generated at different times. The lack of reproducibility has limited their wide use in epidemiological investigations.

Repetitive Extragenic Palindromic PCR (REP-PCR)

Both eukaryotic and prokaryotic species contain highly conserved repetitive DNA sequences throughout the genome. Repetitive extragenic palindromic (REP) DNA sequences were firstly identified in *Escherichia coli* and *Salmonella typhimurium* [65]. The REP sequence is about 40 base pairs long. There are approximately 500–1000 copies of REP sequences interspersed on *E. coli* chromosome DNA, accounting for about 1% of the genome. These sequences were subsequently identified and characterized in many bacterial species [65]. Primers targeting the repetitive sequences allow amplification of DNA regions between the two repetitive DNA sequences in

close proximity and create DNA banding profiles that can be visualized by agarose gel or capillary electrophoresis. A semiautomated REP-PCR system known as DiversiLab was designed with an attempt to offer a standardized and reproducible REP-PCR system to allow interlaboratory comparison. Comparison between REP-PCR and other typing methods such as MLST and pulsed-field gel electrophoresis (PFGE) indicates REP-PCR is at least as discriminative as PFGE for distinguishing vancomycin-resistant *Enterococcus faecium* (VRE) strains and more discriminative for characterizing outbreak strains [66]. So far, REP-PCR has been used to subtype strains of many bacterial species such as *Acinetobacter baumannii*, *Campylobacter* spp., *S. aureus*, and *S. enterica* and compared with MLST analysis [67–70].

PULSED-FIELD GEL ELECTROPHORESIS (PFGE)

PFGE is a DNA fingerprinting method which discriminates bacterial strains based on their genomic DNA pattern generated by digestion with a restriction enzyme, which recognizes specific DNA sequences in the genome and cleaves bacterial genomic DNA into a small number of large fragments (10–45 base pairs) of different sizes. This method involves embedding and lysing bacteria cells in agarose gel blocks, digesting genomic DNA using a specific restriction endonuclease, and size-fractionating the resultant DNA fragments using alternating electric fields. The resulting pattern of distinct DNA bands on agarose gel is referred to as ‘DNA fingerprint’ or ‘PFGE pattern’, which is determined by the number and location of restriction sites within a bacterial genome (Fig. 29.1). While uni-directional conventional gel electrophoresis can only separate DNA fragments up to about 20 kb, PFGE is able to resolve large DNA fragments up to 10 Mb. Different types of PFGE units are currently available including contour-clamped homogeneous electric field (CHEF) [71,72], transverse alternating field electrophoresis (TAFE) [73] and its relative ST/RIDEtm (Stratagene), and rotating gel electrophoresis (RGE) [74,75], with CHEF being the most common one.

PFGE was first described by Schwarz et al. [76] for analysing intact yeast chromosome-sized DNA. It has been successfully expanded to study molecular epidemiology of a number of bacterial pathogens and is considered a standard genetic typing method for many bacterial pathogens. PFGE plays vital roles in outbreak investigations of foodborne diseases. Comparison of PFGE patterns of various bacterial strains provides an important clue as to whether or not these strains are genetically related. Strains with identical PFGE patterns are clonal and probably originated from the same ancestor.

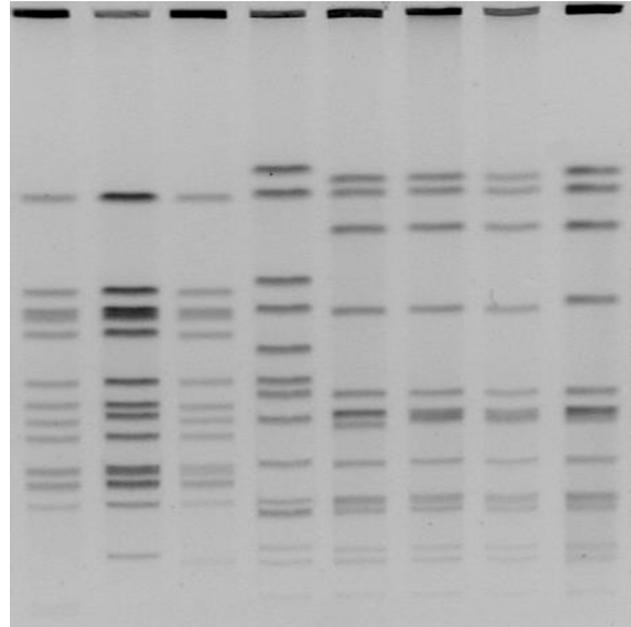


FIGURE 29.1 Pulsed-field gel electrophoresis. Genomic DNA from *Neisseria meningitidis* isolates was digested with NheI and separated by PFGE. Strains with identical or different PFGE patterns are shown as examples.

Conversely, strains with distinct PFGE patterns are not genetically related. Point source outbreak strains are usually very clonal and produce indistinguishable PFGE patterns. However, during an outbreak, genetic variations may occur in strains that undergo frequent horizontal gene transfer and cause a slight change in PFGE patterns. Strains with distinguishable PFGE patterns should be analysed in conjunction with the epidemiological data because they may be epidemiologically linked.

To monitor foodborne diseases worldwide and understand the global molecular epidemiology of these diseases, an international laboratory network known as PulseNet International has been created (www.pulsenetinternational.org). Participating laboratories are from Africa, Asia Pacific, Canada, Europe, Latin America, the Caribbean, the Middle East and the USA. This network allows laboratory researchers to rapidly compare the PFGE patterns of interested strains with existing patterns in the database and determine whether the strains are linked to the outbreak strains. It facilitates intra- or interlaboratory comparison of PFGE patterns of strains from different time and geographic locations. The national PFGE database in the United States is hosted and managed at the Centers for Disease Control and Prevention. Certified participants have access to the database and are able to enter PFGE patterns into the database and analyse them using Bionumerics software as long as standard PFGE protocols are strictly followed.

TABLE 29.2 Software Packages for Multiple Sequence Alignment and Phylogenetic Analysis

Program	Website	Comment
Multiple Sequence Alignment		
Clustal [79]	http://www.clustal.org/	One of the first and most widely used alignment tools, accurate and standard
MUSCLE [80]	http://www.drive5.com/muscle/	Relatively new tool, widely adopted and with exceptional performance
MUMmer [81]	http://mummer.sourceforge.net/	Whole-genome alignment tool
Phylogenetic Analysis Software		
MEGA [82]	http://www.megasoftware.net/	Most highly recommended package, extremely useful, multiple methods implemented, thorough documentation, alignment tool and editor, excellent graphical interface
PHYLIP [77]	http://evolution.gs.washington.edu/phylip.html	Oldest distributed package, wide utility with multiple methods, powerful but not very user friendly
PAUP [83]	http://paup.csit.fsu.edu/	Widely used, emphasis on parsimony but also includes other methods, not free
MrBayes [84]	http://mrbayes.sourceforge.net/	Bayesian inference, highly accurate and widely used, very sensitive to user options
SplitsTree [85]	http://www.splitstree.org/	Method for reconstructing reticulate trees or phylogenetic networks

PHYLOGENETIC TREE RECONSTRUCTION METHODS

Phylogenetic tree reconstruction is a powerful and visually intuitive approach for inferring evolutionary relationships between microbial sequences [77,78]. Continued advances in sequencing technology, along with the growing reliance on sequence-based methods for molecular typing, ensure that the phylogenetic approach will become an increasingly important part of molecular epidemiology studies. There are a number of conceptually distinct methodologies used to reconstruct phylogenetic trees using sequence data along with numerous phylogenetic analysis software packages. The phylogenetic literature is full of debates regarding which of these methods is the best, and there exist vigorously entrenched camps in favour of one method or another method over the others. However, when applied carefully to a reliable data set (i.e. a correct multiple sequence alignment), any of the widely used methods for phylogenetic reconstruction should prove to be largely accurate for inferring evolutionary relationships. More to the point, a robust phylogenetic relationship signal should be present irrespective of the method of reconstruction that is employed. As such, agreement between multiple methods can be taken as a measure of support for an inferred evolutionary relationship of interest, and therefore the use of multiple methods of reconstruction, where appropriate to the data being used, is recommended.

Accordingly, an understanding of the different classes of phylogenetic reconstruction methods is essential for accurate phylogenetic-based molecular typing.

A critical aside relates to the importance of multiple sequence alignment as a prelude to phylogenetic analysis. The adage of garbage in, garbage out rings especially true when it comes to phylogenetic tree reconstruction. The most rigorous methods for phylogenetic reconstruction will not be able to reconstruct accurate evolutionary relationships if they are applied to unreliable multiple sequence alignments. This problem may be less acute with respect to molecular typing since the sequences being analysed are typically highly related and thus easily aligned. Nevertheless, great care should be taken to ensure that the alignments used for phylogenetic reconstruction are accurate. This includes use of the most reliable and up-to-date alignment software packages (Table 29.2) [79–81] along with mandatory visual inspection, and refinement if needed, of any alignment that is to be used in phylogenetic reconstruction.

CLASSES OF PHYLOGENETIC METHODS

There are several different classes of methods that are used to reconstruct phylogenetic trees. The most classic division of phylogenetic methods is between (i) parsimony-based methods, (ii) distance-based methods, and (iii) maximum

likelihood methods. As of late, a substantial amount of effort has been dedicated to the development of additional statistical-based methods for phylogenetic reconstruction such as the Bayesian approach.

Parsimony methods reconstruct a phylogenetic tree that minimizes the number of residue changes between sequences [83]. This method has the advantage that it is based on an intuitive and plausible model of evolution, and it can also assign substitutions to specific branches of the tree. Parsimony methods struggle with so-called homoplasies, or independent changes to the same residue based on convergence or parallelism (i.e. not due to shared common ancestry). Homoplasies may be particularly prevalent among the DNA sequences that are often used in molecular typing, since there are only four distinct DNA sequence residues, and they can lead to the reconstruction of numerous equally parsimonious trees.

Distance-based methods reconstruct phylogenetic trees based on a pairwise distance matrix computed for all taxa (sequences) under consideration [78]. This approach is generally reliable and has the advantage of an intuitive and logical foundation; the most closely related sequences are separated by the shortest distances and vice versa. Accordingly, distance-based approaches, such as the neighbour-joining method [86] and the minimum-evolution methods [87], are most widely used in molecular epidemiology. It should be noted that distance-based approaches can be extremely reliant on the model of substitution used to infer pairwise sequences distances. This concern is mitigated by the fact that molecular typing applications typically analyse closely related sequences, wherein different substitution models will yield similar distances, but it is still important to use the most appropriate substitution model and/or to use multiple models and compare results between models.

Maximum likelihood phylogenetic reconstruction, along with its statistical cousin the Bayesian approach, formally represent the most rigorous and accurate method for reconstructing phylogenetic trees [77]. The maximum likelihood method chooses a phylogenetic tree that maximizes the likelihood of observing the underlying sequence data given a specific substitution model. Maximum likelihood allows for much more flexibility than other approaches, such as varying rates of evolution across sites and lineages, and as such is particularly amenable to addressing very specific evolutionary questions related to sequence change along trees, such as recombination, selection, etc. But maximum likelihood is also the method that requires the most nuanced user input and manipulation, particularly with respect to the sequence substitution model and the tree search parameters. Thus, the maximum likelihood approach may be beyond the scope of many more straightforward molecular typing applications and is

often better suited to teasing apart more distant evolutionary relationships.

Choosing amongst these different approaches can be daunting and therefore an agnostic approach to molecular typing that employs multiple phylogenetic methods is recommended here. Indeed, the use of more than one method of phylogenetic reconstruction can be a good way to assess the support for evolutionary relationships that are of particular interest. Phylogenetic groupings that show up in trees reconstructed with different methods can be considered to be particularly reliable. The ready availability of phylogenetic analysis software packages, their ease of use, and their speed make the multiple-method approach both practical and feasible.

ASSESSING SUPPORT FOR PHYLOGENETIC RELATIONSHIPS

There are additional methods that can be used to assess the support for the evolutionary relationships inferred using the phylogenetic approaches described above. The most commonly employed method is the bootstrap, which relies on generating numerous phylogenies from alignments that are randomly resampled (with replacement) from the original sequence alignment. The percentage of time that any particular internal branch (or equivalently any internal node) shows up amongst the set of trees reconstructed from the resampled alignment is taken as the bootstrap support measure for the group delineated by that branch (node). Since it relies on resampling, the bootstrap is a non-parametric method and thus the bootstrap support values that are obtained do not have a specific statistical interpretation. It has been estimated that a bootstrap value of ≥ 70 is roughly equivalent to a confidence interval of 95%. Bootstrap values ≥ 90 are considered to indicate extremely reliable phylogenetic relationships.

RETICULATE TREE RECONSTRUCTION

It is important to note that phylogenies, and their underlying graphic structures, represent a specific and limited model of the evolutionary process. Phylogenies are structured as directed acyclic graphs, or bifurcating trees, which therefore imply a strictly diversifying evolutionary process amongst the microbes being typed. According to the phylogenetic model, microbial strains would genetically diversify and the resulting distinct sequence pools would never be brought back into contact. In fact, it is well known that microbial evolution often deviates from a strictly diversifying mode via recombination and horizontal transfer [85]. This type of evolution, whereby sequences that have been separated over evolutionary time are later reintroduced to the same genetic background, is considered to be reticulate, and there are

phylogenetic methods developed to capture and represent this reticulate evolutionary process (SplitsTree in Table 29.2). The use of such reticulate methods may yield a more nuanced and accurate view of the evolutionary relationships among a group of microbial strains.

GENOME-WIDE PHYLOGENETIC RECONSTRUCTION

As whole-genome sequencing becomes increasingly rapid and cost-effective, genome-wide approaches to molecular typing and phylogenetic reconstruction will become more prevalent and necessary. Genome-wide phylogenetic reconstruction typically proceeds using the same general approaches described above, albeit using much larger data sets consisting of whole-genome sequence alignments or concatenated alignments of multiple proteins. Whole-genome sequence alignments are typically used to infer inter-genome distances, but these distances may be confounded by issues of horizontal transfer and/or lineage-specific gene loss and duplication. To avoid this problem, genome-wide phylogenetic reconstruction is often conducted using concatenated alignments of so-called ‘core genes’ that are present amongst all strains being analysed and considered to be more resistant to horizontal transfer. Such a set of ‘core genes’ can be thought of as a much larger set of genes that is analogous to the dispersed housekeeping genes used in MLST. An alternative approach to the use of concatenated alignments is the ‘super-tree’ approach whereby individual trees are reconstructed from single gene alignments and later merged to form a single overall genome tree. This method is more sensitive to slight differences between individual genes and thus perhaps slightly less reliable when applied to genome-wide molecular typing.

PHYLOGENETIC SOFTWARE PACKAGES

The most widely used and readily available software packages that can be used for sequence-based phylogenetic reconstruction are summarized in Table 29.2. This list reflects our own bias and familiarity and is by no means exhaustive; the most comprehensive list of phylogeny programs is maintained at <http://evolution.genetics.washington.edu/phylip/software.html>.

PHYLOGENETIC ANALYSIS OF BACTERIAL PATHOGENS

Tracing *Neisseria meningitidis*

N. meningitidis is a human commensal bacterium and also one of the common pathogens causing bacterial meningitis. This organism is classified into 12 serogroups based

on the structure and chemical composition of cell-surface-associated capsular polysaccharide. Only six serogroups (A, B, C, W, X and Y) are associated with most invasive disease. Meningococcal disease remains a significant public health concern due to its high morbidity and mortality rate worldwide. The highest incidence (up to 1000 cases per 100 000 population) has been found in the meningitis belt of sub-Saharan Africa, where large devastating epidemics occurred periodically, with serogroup A as the major cause [88,89]. Meningococcal meningitis epidemics were also reported in many other countries in North America, Asia, Europe and South America. With the implementation of meningococcal vaccines, there has been a shift in the global epidemiology of meningococcal disease [90,91].

Phylogenetic Typing of *N. meningitidis*

The *N. meningitidis* population is antigenically and genetically diverse. The diversity results mostly from frequent horizontal gene transfer events, and sometimes from vaccine-induced immune selection, which may provide an additional driving force for diversity of surface-exposed antigens [92,93]. Several typing schemes have been used to measure meningococcal genetic and antigenic diversity [94]. Multilocus enzyme electrophoresis (MLEE) had long been considered the gold standard for phylogenetic typing of meningococci before the development of MLST. Strains of meningococci were grouped into different subgroups/lineages/complexes (e.g. subgroups I–VIII; lineage 3; ET-37 complex) and electrophoretic types such as ET-5, ET-37 and ET-15 [95–97] based on the genetic similarity of multiple loci. Strains that were associated with outbreaks, epidemics or endemics were assigned as different hypervirulent lineages. Since it was proposed, MLST has become a general approach for the studies of molecular epidemiology and population biology of *N. meningitidis*. Major clonal complexes (CC) such as CC11, CC32, CC41/44, CC23 and CC181 are recognized to have a close association with invasive disease. Correlation between MLEE and MLST designations has been established. In general, MLST and MLEE are congruent in identifying hypervirulent lineages with only a few exceptions. For example, MLEE subgroup I/II is corresponding to clonal complex 5 (CC1) by MLST, and subgroups III and IV to CC5 and CC4, respectively (Pubmlst.org). However, MLST was not able to distinguish strains between subgroups I and II, V and VII, IV-1 and IV-2, or III and VIII. As the subgroups were defined by both MLEE and the outer-membrane protein-encoding genes, the genetic differences between the paired subgroups may result from outer-membrane protein genes. Subgroup III merged with subgroup VIII when OMP genes were excluded from the cluster analysis [97]. Other methods

such as SNP and whole-genome-based typing are used to detect comprehensive genetic variations of this organism [98,99]. However, these methods are still in the early stage of development. Clearly defined typing schemes are still lacking or require further validation. Surface antigens porin A (PorA) and ferric enterobactin transport (FetA) are major components of some outer-membrane vesicle (OMV)-based vaccines. PorA and FetA sequencing methods detect antigenic variations that occur within the surface-exposed variable regions of these proteins. A typing scheme combining the above molecular methods has provided added value for assessment of genetic microevolution and vaccine impact on meningococcal genetic and antigenic structure [100]. Although a direct impact of meningococcal vaccine on molecular epidemiology of *N. meningitidis* remains to be further evaluated, *N. meningitidis* has evolved in the vaccine era, which leads to changes in population structure of this pathogen.

Clonality of Meningococcal Population Structure

N. meningitidis serogroup A has been historically linked to the majority of the large meningitis epidemics worldwide since it was first isolated and identified [101]. Early phylogenetic analyses of *N. meningitidis* serogroup A were mostly done by MLEE, which have revealed that serogroup A meningococcal strains possess a clonal population structure. Using genes coding for seven enzymes and two outer-membrane proteins, Olyhoek et al. detected 34 electrophoretic types among the 423 epidemic- or endemic-causing isolates that were collected from 38 countries between 1915 and 1983 [101,102]. The 34 ETs formed four phylogenetic groups using cluster analysis, which were designated as subgroups I–IV. Each subgroup contains four or more clones (I: 1–9; II: 1–4; III: 1–4; and VI: 1–4). A different study characterized 290 strains chosen from all major meningitis epidemics worldwide from 1915 to 1991 using genes encoding 15 cytoplasmic enzymes and four outer-membrane proteins. The study defined 84 ETs belonging to nine subgroups. Both studies demonstrated that epidemics caused by serogroup A meningococci were highly associated with certain clones, the most important of which were subgroups I/CC1, III/CC5 and VI-1/CC4. Strains of these subgroups are believed to come from the same ancestor [97,102,103]. The clonal population structure of *N. meningitidis* serogroup A strains were also proved by other molecular typing schemes including MLST and random amplified polymorphic DNA (RAPD) [101].

Clonal structure exists among strains of other *N. meningitidis* serogroups as well. Hypervirulent lineages or clones have been identified to cause epidemics, local outbreaks or endemics. For example, CC11/ET-37

complex, CC23/cluster A3, CC32/ET-5 complex and CC41/44/lineage 3 are the predominant clones that cause invasive meningococcal disease. CC11/ET-37 complex is mainly associated with serogroup C as well as serogroup W. CC23/cluster A3 accounts for the majority of serogroup Y strains worldwide. Serogroup B is genetically more diverse than other serogroups, with CC32/ET-5 complex and CC41/44/lineage 3 being the common clones followed by CC18 and CC269 [91].

Temporal and Spatial Shift in Molecular Epidemiology of *N. meningitidis*

N. meningitidis is an exemplary pathogen that exhibits spatial and temporal fluidity in epidemiology. CC1/subgroup I was responsible for a number of epidemics in Africa between the 1960s and 1970s, including the global pandemics that started in 1967 in the Mediterranean and North Africa and then spread throughout West Africa during 1968–72. The epidemics in Niger during 1960–63 and in Burkina Faso during 1969–73 were also associated with CC4/subgroup IV. The epidemic waves in the early 1980s in Africa were caused exclusively by CC4/subgroup IV [97,102]. Strains of CC5/subgroup III have the potential to cause large meningococcal meningitis epidemics. Two historical pandemics were associated with CC5/subgroup III strains. One pandemic wave affected China, Finland, Moscow, Norway and Brazil in the mid-1960s. During the 1980s, the same clone re-emerged and caused a second meningitis pandemic wave that began in China and Nepal and then spread to Mecca, Saudi Arabia during the Hajj pilgrimage in 1987. In 1988, CC5/subgroup III epidemics occurred and progressed from Eastern Africa to Central and Western Africa [89,97,104]. The largest epidemic wave caused by CC5/subgroup III occurred in 1996 and primarily affected Nigeria and Burkina Faso. Although CC5/subgroup III remained the predominant cause of meningococcal epidemics in Africa [88], strains of CC5/subgroup III have evolved during this period; new sequence types (STs) appeared and the old STs were being replaced. During 1988–2003, both ST-5 and ST-7 were detected in Africa, with ST-7 being the prevalent sequence type among serogroup A strains [105]. But in the period of 2004–10, ST-5 was replaced by ST-7; ST-2859 that first appeared in Burkina Faso in 2003 dominated within CC5 from a number of countries. Three new sequence types (ST-5788, -6968, -8639) emerged and accounted for a very small proportion of serogroup A strains during this period. The new STs are all single-locus variants. Allelic evolution of the new STs is shown in Table 29.3.

In 2010, a serogroup A conjugate vaccine (MenAfriVac) was first introduced in Burkina Faso, Niger

TABLE 29.3 Single Locus Variants within Clonal Complex CC5

Clonal Complex	Sequence Type	<i>abcZ</i>	<i>adk</i>	<i>aroE</i>	<i>fumC</i>	<i>gdh</i>	<i>pdhC</i>	<i>pgm</i>
CC5	ST-5	1	1	2	1	3	2	3
CC5	ST-7	1	1	2	1	3	2	19
CC5	ST-2859	1	3	2	1	3	2	19
CC5	ST-5788	1	1	437	1	3	2	19
CC5	ST-6968	1	3	2	1	3	456	19
CC5	ST-8639	1	1	2	1	3	2	578

and Mali and will be gradually implemented in all African countries by 2016. Significant changes in meningococcal molecular epidemiology are expected in the post-vaccine era. The vaccinated countries have already witnessed epidemiological changes, with a significant reduction in serogroup A cases and an increase in the proportion of serogroup W strains [106,107]. Serogroup W has been found to circulate in Europe, the United States and Africa since the 1980s and generally causes sporadic cases or small-scale outbreaks. The first large serogroup W outbreak occurred during the Hajj pilgrimage in Saudi Arabia in 2000 and was caused by a strain of CC11/ET-37 complex with a PorA type of P1.5,2 designated as the Hajj clone. CC11/ET-37 complex is usually associated with serogroup C. It was suspected that capsule switching occurred between serogroups W and C, but when and in which direction is not known [31,108].

The Hajj outbreak marked the beginning of intercontinental spread of serogroup W. This strain was rapidly spread by the Hajj pilgrims throughout European, Asian, Middle Eastern and North African countries and the United States. A large outbreak involving nine European countries (the UK, France, the Netherlands, Germany, Finland, Sweden, Belgium, Switzerland and Norway) was reported immediately following the Hajj outbreak in 2000 [31,109]. In 2002, a serogroup W epidemic was reported in Burkina Faso, which affected 12 000 people and caused 1400 deaths. A phylogenetic analysis indicated the Hajj outbreak-associated strains collected from Saudi Arabia, France, Singapore, Finland and the United States in 2000 had identical PFGE pattern (H46N06.0040), 16S type (type 31) and PorA type (P1.5,2), and belonged to the CC11 (ST-11)/ET-37 complex (ET-27). However, the strains from 1970–2000 that were not epidemiologically linked to the Hajj outbreak showed diverse genetic backgrounds. None of these strains belonged to 16S type 31. Some of the non-Hajj-linked strains were genetically identical to the Hajj clone by PFGE, MLEE and PorA typing, yet had a different 16S type, type 13, which differs from type 31 in three nucleotides. These lines of

evidence suggested that the Hajj-related clone was circulating in different regions before 2000, and the Hajj outbreak probably amplified the global transmission of this clone. Due to the recombining nature of *N. meningitidis*, genetic variation occurred during the transmission as indicated by the difference in 16S type among these strains [31]. The epidemic strain in Burkina Faso was also associated with the Hajj clone. CC11/ET-37 complex has been present in Africa since at least 1993 and continues to spread in African countries after the 2002 Burkina Faso epidemic [105,110]. CC11 (ST-11)/ET-37 complex remains to be the prevalent genotype. New STs (ST-5779 and ST-8637) and clonal complex (CC175) emerged between 2002 and 2010 [88,105]. Figure 29.2 illustrates the genetic relatedness of some serogroup W strains collected from different regions (Wang and Mayer, unpublished data). ET-27 serogroup W strains are closely related to ET-27 serogroup C strains by PFGE analysis.

Serogroups B and C are universally present on most continents. Both serogroups are clonal in structure with serogroup B strains being more heterogeneous. Serogroup C is less frequently associated with epidemics than serogroup A. CC11 (ST-11)/ET-37 complex was responsible for the epidemics in Brazil in the 1970s and has persisted in Brazil since then [111,112]. A new ET-11 complex emerged in Sao Paulo and other Brazilian states in late 1990s and CC103 from 2002 onwards [111,113]. Starting from the 1990s, there has been an increase in CC11/ET-37 complex serogroup C disease in North America and Europe. This clone has been responsible for the majority of the outbreaks and endemics before the implementation of serogroup C vaccines in North America and Europe, and remains a major cause of meningococcal disease in the United States in the 21st century, followed by CC103 [90,113,114].

Being associated with both epidemic and endemic situations, serogroup B has raised increasing concerns in Europe and the Americas since the 1970s. Epidemics caused by serogroup B CC41/44/lineage 3 have been reported in the Netherlands between 1965 and 1966 and

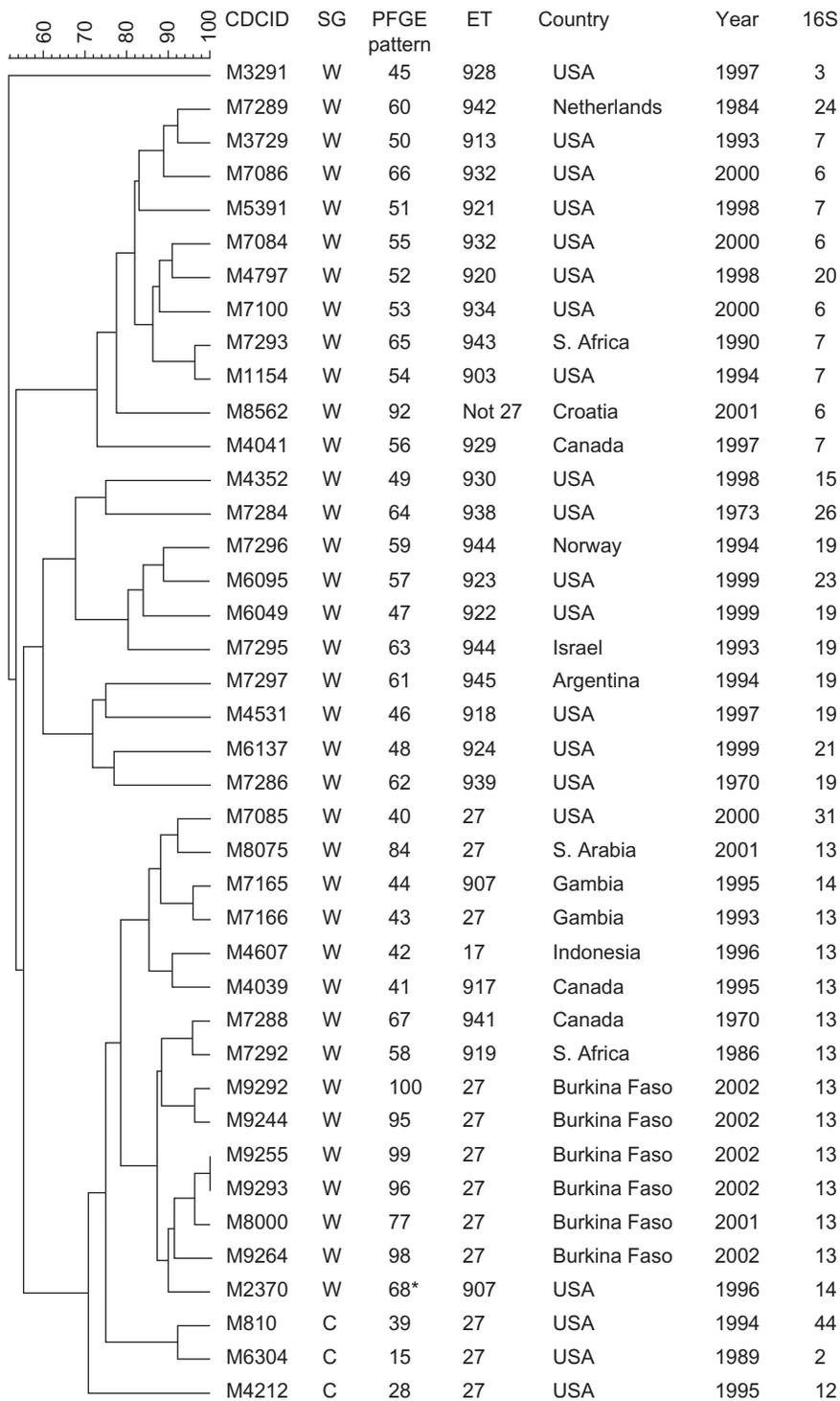


FIGURE 29.2 Genetic diversity of serogroup W strains from different regions. The phylogenetic tree is based on PFGE patterns. The ET types and 16S types of these strains are also indicated. ET-27 strains of serogroups W and C clustered within the same group.

in Belgium during 1969–72. Strains within this clone were also responsible for the increased meningococcal cases in Italy and France during this period. CC32/ET-5 complex first appeared in Norway causing an epidemic in 1975. All members within this clone from different regions in Europe are closely related, suggesting its spread from Norway to other European countries [96].

In the 1980s, Cuba experienced an epidemic caused by CC32/ET-5 complex, which subsequently spread to Sao Paulo, Brazil in the late 1980s. Both CC41/44/lineage 3 and CC32/ET-5 complex are common in the United States, with CC32/ET-5 complex being the major cause of a prolonged serogroup B hyperendemic in Oregon state since 1993 [115].

Phylogeny for Classification of *H. influenzae* and *H. haemolyticus*

H. influenzae and *H. haemolyticus* are human commensals colonizing in the pharyngeal cavity. Encapsulated or typeable *H. influenzae* produces one of the six types of capsule (a–f) with serotype b being the major cause of life-threatening invasive disease such as meningitis and bacteraemia. Unencapsulated or non-typeable *H. influenzae* does not produce capsule and is a common cause of non-invasive diseases including otitis media, bronchitis and pneumonia worldwide [116]. *H. haemolyticus* rarely causes disease but was recently found to be responsible for a small number of invasive disease cases in the United States. These *H. haemolyticus* cases have been historically overlooked and misidentified as non-typeable *H. influenzae* due to the high similarity between the two organisms in genetic composition and physiology [27,28].

As a naturally transformable species, *H. influenzae* undergoes frequent horizontal gene exchange and is highly genetically diverse [117,118]. The presence of competence genes in *H. haemolyticus* strains suggests this organism is probably naturally transformable as well [119]. Because both organisms share the same biological niche, the high degree of genetic similarity may result from DNA exchange between the two organisms through natural transformation. Haemolytic *H. haemolyticus* produces a clear zone on horse blood agar plate, which differentiates it from *H. influenzae*. However, it remains a challenge to discriminate *H. influenzae* from non-haemolytic *H. haemolyticus* [120,121].

While DNA–DNA hybridization remains the gold standard for identifying bacterial species, 16S rRNA gene sequencing has been used to infer bacterial phylogeny at the species and genus level, and assists with discrimination between closely related bacterial species such as *H. influenzae* and *H. haemolyticus* [21,28,121,122]. Strains of *H. influenzae* and *H. haemolyticus* can basically form two distinct groups on a dendrogram constructed on 16S rRNA gene sequence with a few exceptions where some *H. haemolyticus* strains cluster with the *H. influenzae* group [120,121]. Studies also indicate that the phylogeny of *Haemophilus* species based on partial sequencing of three (*adk*, *pgi* and *recA*) of the seven MLST loci (*adk*, *atpG*, *frdB*, *fucK*, *mdh*, *pgi* and *recA*) and a gene coding for translation initiation factor 2 (*infB*) was in agreement with DNA–DNA hybridization data [123–125]. A combined scheme using concatenated sequences of three (*adk*, *pgi*, and *recA*) or six (*adk*, *atpG*, *frdB*, *mdh*, *pgi* and *recA*) MLST loci, *infB* and 16S rRNA gene was developed in an attempt to improve discrimination between *H. influenzae* and *H. haemolyticus* [120,121]. Phylogeny inferred from multilocus gene sequences of *infB* and MLST loci defined closely related species *H. influenzae*, *H. aegyptius*

(also considered as *H. influenzae*) and *H. haemolyticus* as one group. As shown in Figure 29.2, ambiguity was observed in using these methods for distinguishing *H. influenzae* and *H. haemolyticus* [28,120]. Strains that exhibit the traits of *H. haemolyticus* group but belong to *H. influenzae* phylogeny are defined as variant or fuzzy strains. Unambiguous classification of these strains may require high discriminatory phylogenetic tool such as genome-based typing. Preliminary genomic analysis of *H. influenzae* and *H. haemolyticus* revealed two distinct groups on a dendrogram of core genes (data not shown). A clearly defined typing scheme remains to be developed for classifying the two closely related species.

Reconstruction of *Yersinia pestis* History by Phylogenetics

Y. pestis is a monomorphic pathogen and causative agent of a human invasive disease known as ‘plague’. The organism is classified into three biovars (Antiqua, Medievalis and Orientalis) based on the sugar metabolism and ability to reduce nitrate. Three historical pandemics caused by *Y. pestis* have been documented since AD 541. The first plague, also known as Justinian’s plague, started in Egypt during AD 541–542 and spread to the Mediterranean regions. The Black Death, the second plague pandemic, initiated from the Caspian Sea and swept across all countries in Europe. A number of epidemic waves of plague continued until 1750. One-third of the European population died during this pandemic. The third plague pandemic originated from south China in the mid-19th century, and spread to Hong Kong and subsequently to Europe, Africa, the Americas and India through marine shipping. Evidence has suggested the three biovars were responsible for the three pandemics, with Antiqua possibly associated with the first pandemic, Medievalis possibly with the second pandemic and Orientalis with the third pandemic [126–128]. Studies have been undertaken to understand the population structure of this pathogen and reconstruct its history.

Phylogenetic analyses based on independent methods (e.g. MLVA, the presence of IS100, SNPs, and ribotyping) indicate that *Y. pestis* is a clonal pathogen that evolved from *Yersinia pseudotuberculosis* shortly before the Justinian’s plague [13,52,129]. Although the two species are highly related by DNA–DNA hybridization and have identical 16S rRNA genes, the fact that they each follow different transmission pathways and differ in virulence suggests they should be considered different species. Phylogeny deduced from sequences of five housekeeping genes (*thrA*, *trpE*, *glnA*, *tmk* and *dmsA*) and a gene (*manB*) involved in lipopolysaccharide biosynthesis supports this classification. *Y. pestis* shared the same

alleles of *thrA*, *glnA* and *manB* with some strains of *Y. pseudotuberculosis*. A phylogenetic tree based on the sequences of the five genes illustrated that strains from both species clustered together, which form a distinct group from *Yersinia enterocolitica* [129]. These studies also illustrated that the three biovars were separated into phylogenetically distinct groups, with Antiqua emerging first, followed by Medievalis and Orientalis. Additional analysis using Sequenom MassArray SNP typing proposed a model for *Y. pestis* migrating from its origin in East Asia to other regions worldwide [54]. *Y. pestis* Antiqua was the first biovar to emerge and originally evolved in China more than 2600 years ago and circulated in surrounding regions such as Mongolia and the former Soviet Union. Then Medievalis diverged from Antiqua more than 545 years ago during the Silk Road period. Strains of Medievalis were found in western China, Kazakhstan and the Caucasus, suggesting Medievalis spread from China to these countries along the Silk Road. The youngest biovar, Orientalis, evolved more than 210 years ago and was subsequently disseminated to multiple countries including Europe, South America, Africa, Southeast Asia, Madagascar and Turkey. The spreading of Orientalis is thought to be associated with an historical event, Zheng He's voyage from China to Africa, which happened at about the same time period [54].

CONCLUDING REMARKS

Both classical and advanced molecular typing schemes are reviewed in this chapter. Exemplar applications on a few extensive phylogenetic studies are also provided. While classical typing schemes remain valuable in many aspects, next-generation DNA sequencing technology has been increasingly used in routine practice in clinical and research microbiology laboratories and represents the future. DNA sequencing technology has played crucial roles in the understanding of disease transmission and investigating molecular epidemiology and population structure of important human pathogens. With the rapid expansion of publicly accessible database and automated bioinformatics tools, clinical laboratories are expected to be soon transformed by implementing next-generation sequencing technology.

Many important discoveries in the 21st century have been achieved through whole-genome sequencing approaches. Although it has been used in many epidemiological investigations and bacterial classification, whole-genome-based typing is still at its early development stage. Genome-based typing schemes remain to be clearly defined and extensively validated. Automated pipelines for genome assembly and annotation, and comparative genome analysis that can be customized for all bacterial pathogens are still under development. However, great

efforts are being made to address these challenges. Whole-genome sequencing will hopefully be integrated into laboratory routine practice in the near future.

ACKNOWLEDGEMENT

The authors thank Susanna Schmink for providing Figures 29.1 and 29.2.

REFERENCES

- [1] Gordis L. Epidemiology. 4th ed. Philadelphia, PA: Elsevier/Saunders; 2008.
- [2] Nelson KE, Williams CM, editors. Infectious Disease Epidemiology: Theory and Practice. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers; 2007.
- [3] Foxman B, Riley L. Molecular epidemiology: focus on infection. *Am J Epidemiol* 2001;153:1135–41.
- [4] Parkhill J, Wren BW. Bacterial epidemiology and biology – lessons from genome sequencing. *Genome Biol* 2011;12:230.
- [5] Baker S, Hanage WP, Holt KE. Navigating the future of bacterial molecular epidemiology. *Curr Opin Microbiol* 2010;13:640–5.
- [6] Foxman B, Zhang L, Koopman JS, Manning SD, Marrs CF. Choosing an appropriate bacterial typing technique for epidemiologic studies. *Epidemiol Perspect Innov* 2005;2:10.
- [7] Lukinmaa S, Nakari UM, Eklund M, Siitonen A. Application of molecular genetic methods in diagnostics and epidemiology of food-borne bacterial pathogens. *APMIS* 2004;112:908–29.
- [8] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;95:3140–5.
- [9] Archer GL, Karchmer AW, Vishniavsky N, Johnston JL. Plasmid-pattern analysis for the differentiation of infecting from noninfecting *Staphylococcus epidermidis*. *J Infect Dis* 1984;149:913–20.
- [10] Singh A, Goering RV, Simjee S, Foley SL, Zervos MJ. Application of molecular techniques to the study of hospital infection. *Clin Microbiol Rev* 2006;19:512–30.
- [11] Grimont F, Grimont PA. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Ann Inst Pasteur Microbiol* 1986;137B:165–75.
- [12] Goering RV. Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. *Infect Genet Evol* 2010;10:866–75.
- [13] Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 2008;62:53–70.
- [14] Hall BG, Ehrlich GD, Hu FZ. Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. *Microbiology* 2010;156:1060–8.
- [15] Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012;158:1005–15.
- [16] Olsen RJ, Long SW, Musser JM. Bacterial genomics in infectious disease and the clinical pathology laboratory. *Arch Pathol Lab Med* 2012;136:1414–22.
- [17] Woese CR. Bacterial evolution. *Microbiol Rev* 1987;51:221–71.

- [18] Brenner DJ, Martin MA, Hoyer BH. Deoxyribonucleic acid homologies among some bacteria. *J Bacteriol* 1967;94:486–7.
- [19] Clarridge III JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004;17:840–62, table of contents.
- [20] Woo PC, Lau S, Teng JL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect* 2008;14:908–34.
- [21] Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 2007;45:2761–4.
- [22] Wisotzkey JD, Jurtshuk Jr P, Fox GE, Deinhard G, Poralla K. Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. *Int J Syst Bacteriol* 1992;42:263–9.
- [23] Bavykin SG, Lysov YP, Zakhariyev V, Kelly JJ, Jackman J, Stahl DA, et al. Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J Clin Microbiol* 2004;42:3711–30.
- [24] Sacchi CT, Alber D, Dull P, Mothershed EA, Whitney AM, Barnett GA, et al. High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. *J Clin Microbiol* 2005;43:3734–42.
- [25] Sacchi CT, Whitney AM, Mayer LW, Morey R, Steigerwalt A, Boras A, et al. Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. *Emerg Infect Dis* 2002;8:1117–23.
- [26] Sacchi CT, Whitney AM, Reeves MW, Mayer LW, Popovic T. Sequence diversity of *Neisseria meningitidis* 16S rRNA genes and use of 16S rRNA gene sequencing as a molecular subtyping tool. *J Clin Microbiol* 2002;40:4520–7.
- [27] Anderson R, Wang X, Briere EC, Katz LS, Cohn AC, Clark TA, et al. *Haemophilus haemolyticus* isolates causing clinical disease. *J Clin Microbiol* 2012;50:2462–5.
- [28] Theodore MJ, Anderson RD, Wang X, Katz LS, Vuong JT, Bell ME, et al. Evaluation of new biomarker genes for differentiating *Haemophilus influenzae* from *Haemophilus haemolyticus*. *J Clin Microbiol* 2012;50:1422–4.
- [29] Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics* 2001;17:1230–1.
- [30] Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* 2003;11:479–87.
- [31] Mayer LW, Reeves MW, Al-Hamdan N, Sacchi CT, Taha MK, Ajello GW, et al. Outbreak of W135 meningococcal disease in 2000: not emergence of a new W135 strain but clonal expansion within the electrophoretic type-37 complex. *J Infect Dis* 2002;185:1596–605.
- [32] Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* 2008;6:431–40.
- [33] Stull TL, LiPuma JJ, Edlind TD. A broad-spectrum probe for molecular epidemiology of bacteria: ribosomal RNA. *J Infect Dis* 1988;157:280–6.
- [34] Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009;462:1056–60.
- [35] Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinf* 2010;11:595.
- [36] Abby S, Daubin V. Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* 2007;15:135–41.
- [37] Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, et al. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci USA* 2006;103:5977–82.
- [38] Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011;331:430–4.
- [39] Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci USA* 2005;102:13950–5.
- [40] Papke RT, Gogarten JP. Ecology. How bacterial lineages emerge. *Science* 2012;336:45–6.
- [41] Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, et al. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol* 2007;8:R103.
- [42] Santos SR, Ochman H. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* 2004;6:754–9.
- [43] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;269:496–512.
- [44] Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012;13:601–12.
- [45] Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, et al. A genomic day in the life of a clinical microbiology laboratory. *J Clin Microbiol* 2013;51:1272–7.
- [46] Beckloff N, Starkenburg S, Freitas T, Chain P. Bacterial genome annotation. *Methods Mol Biol* 2012;881:471–503.
- [47] Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 2012;40:D115–22.
- [48] Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. *N Engl J Med* 2011;364:33–42.
- [49] Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *MBio* 2011;2:e00157–11.
- [50] Piarroux R, Barrais R, Faucher B, Haus R, Piarroux M, Gaudart J, et al. Understanding the cholera epidemic, Haiti. *Emerg Infect Dis* 2011;17:1161–8.
- [51] Foley SL, Zhao S, Walker RD. Comparison of molecular typing methods for the differentiation of *Salmonella* foodborne pathogens. *Foodborne Pathog Dis* 2007;4:253–76.
- [52] Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, et al. Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA* 2004;101:17837–42.
- [53] Holt KE, Baker S, Dongol S, Basnyat B, Adhikari N, Thorson S, et al. High-throughput bacterial SNP typing identifies distinct clusters of *Salmonella* Typhi causing typhoid in Nepalese children. *BMC Infect Dis* 2010;10:144.

- [54] Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 2010;42:1140–3.
- [55] Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 2009;4:e7815.
- [56] Lindstedt BA. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 2005;26:2567–82.
- [57] Grissa I, Bouchon P, Pourcel C, Vergnaud G. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. *Biochimie* 2008;90:660–8.
- [58] van Belkum A, Scherer S, van Leeuwen W, Willemse D, van Alphen L, Verbrugh H. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. *Infect Immun* 1997;65:5017–27.
- [59] Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, et al. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–36.
- [60] Klevytska AM, Price LB, Schupp JM, Worsham PL, Wong J, Keim P. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* 2001;39:3179–85.
- [61] Kruy SL, van Cuyck H, Koeck JL. Multilocus variable number tandem repeat analysis for *Salmonella enterica* subspecies. *Eur J Clin Microbiol Infect Dis* 2011;30:465–73.
- [62] Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramière V, et al. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* 2001;1:2.
- [63] Roussel S, Felix B, Grant K, Dao TT, Brisabois A, Amar C. Fluorescence amplified fragment length polymorphism compared to pulsed field gel electrophoresis for *Listeria monocytogenes* subtyping. *BMC Microbiol* 2013;13:14.
- [64] Mueller UG, Wolfenbarger LL. AFLP genotyping and fingerprinting. *Trends Ecol Evol* 1999;14:389–94.
- [65] Lupski JR, Weinstock GM. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* 1992;174:4525–9.
- [66] Bourdon N, Lemire A, Fines-Guyon M, Auzou M, Perichon B, Courvalin P, et al. Comparison of four methods, including semi-automated rep-PCR, for the typing of vancomycin-resistant *Enterococcus faecium*. *J Microbiol Methods* 2011;84:74–80.
- [67] Behringer M, Miller WG, Oyarzabal OA. Typing of *Campylobacter jejuni* and *Campylobacter coli* isolated from live broilers and retail broiler meat by *flaA*-RFLP, MLST, PFGE and REP-PCR. *J Microbiol Methods* 2011;84:194–201.
- [68] Ben-Darif E, De Pinna E, Threlfall EJ, Bolton FJ, Upton M, Fox AJ. Comparison of a semi-automated rep-PCR system and multilocus sequence typing for differentiation of *Salmonella enterica* isolates. *J Microbiol Methods* 2010;81:11–6.
- [69] Higgins PG, Janssen K, Fresen MM, Wisplinghoff H, Seifert H. Molecular epidemiology of *Acinetobacter baumannii* bloodstream isolates obtained in the United States from 1995 to 2004 using rep-PCR and multilocus sequence typing. *J Clin Microbiol* 2012;50:3493–500.
- [70] Te Witt R, Kanhai V, van Leeuwen WB. Comparison of the DiversiLab system, pulsed-field gel electrophoresis and multi-locus sequence typing for the characterization of epidemic reference MRSA strains. *J Microbiol Methods* 2009;77:130–3.
- [71] Chu G. Pulsed-field electrophoresis: theory and practice. *Methods* 1990;1:129–42.
- [72] Chu G, Vollrath D, Davis RW. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* 1986;234:1582–5.
- [73] Gardiner K, Laas W, Patterson D. Fractionation of large mammalian DNA restriction fragments using vertical pulsed-field gradient gel electrophoresis. *Somat Cell Mol Genet* 1986;12:185–95.
- [74] Serwer P, Hayes SJ. A new mode of rotating gel electrophoresis for fractionating linear and circular duplex DNA: the effects of electrophoresis during the gel's rotation. *Appl Theor Electrophor* 1989;1:95–8.
- [75] Southern EM, Anand R, Brown WR, Fletcher DS. A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res* 1987;15:5925–43.
- [76] Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984;37:67–75.
- [77] Felsenstein J. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates; 2004.
- [78] Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
- [79] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23:2947–8.
- [80] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–7.
- [81] Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol* 2004;5:R12.
- [82] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–9.
- [83] Swofford DL. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sunderland, MA: Sinauer Associates; 2003.
- [84] Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17:754–5, table.
- [85] Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–67.
- [86] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–25.
- [87] Rzhetsky A, Nei M. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* 1993;10:1073–95.
- [88] Caugant DA, Kristiansen PA, Wang X, Mayer LW, Taha MK, Ouedraogo R, et al. Molecular characterization of invasive meningococcal isolates from countries in the African meningitis belt before introduction of a serogroup A conjugate vaccine. *PLoS One* 2012;7:e46019.
- [89] Caugant DA, Nicolas P. Molecular surveillance of meningococcal meningitis in Africa. *Vaccine* 2007;25(Suppl. 1):A8–11.

- [90] Harrison LH, Trotter CL, Ramsay ME. Global epidemiology of meningococcal disease. *Vaccine* 2009;27(Suppl. 2):B51–63.
- [91] Racloz VN, Luiz SJ. The elusive meningococcal meningitis serogroup: a systematic review of serogroup B epidemiology. *BMC Infect Dis* 2010;10:175.
- [92] Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface* 2007;4:787–802.
- [93] Maiden MC. Population genomics: diversity and virulence in the *Neisseria*. *Curr Opin Microbiol* 2008;11:467–71.
- [94] Diggle MA, Clarke SC. Molecular methods for the detection and characterization of *Neisseria meningitidis*. *Expert Rev Mol Diagn* 2006;6:79–87.
- [95] Caugant DA, Bovre K, Gaustad P, Bryn K, Holten E, Hoiby EA, et al. Multilocus genotypes determined by enzyme electrophoresis of *Neisseria meningitidis* isolated from patients with systemic disease and from healthy carriers. *J Gen Microbiol* 1986;132:641–52.
- [96] Caugant DA, Froholm LO, Bovre K, Holten E, Frasch CE, Mocca LF, et al. Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. *Proc Natl Acad Sci USA* 1986;83:4927–31.
- [97] Wang JF, Caugant DA, Li X, Hu X, Poolman JT, Crowe BA, et al. Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China. *Infect Immun* 1992;60:5267–82.
- [98] Katz LS, Sharma NV, Harcourt BH, Thomas JD, Wang X, Mayer LW, et al. Using single-nucleotide polymorphisms to discriminate disease-associated from carried genomes of *Neisseria meningitidis*. *J Bacteriol* 2011;193:3633–41.
- [99] Krauland MG, Dunning Hotopp JC, Riley DR, Daugherty SC, Marsh JW, Messonnier NE, et al. Whole-genome sequencing to investigate the emergence of clonal complex 23 *Neisseria meningitidis* serogroup Y disease in the United States. *PLoS One* 2012;7:e35699.
- [100] Brehony C, Jolley KA, Maiden MC. Multilocus sequence typing for global surveillance of meningococcal disease. *FEMS Microbiol Rev* 2007;31:15–26.
- [101] Bart A, Barnabe C, Achtman M, Dankert J, van der Ende A, Tibayrenc M. The population structure of *Neisseria meningitidis* serogroup A fits the predictions for clonality. *Infect Genet Evol* 2001;1:117–22.
- [102] Olyhoek T, Crowe BA, Achtman M. Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983. *Rev Infect Dis* 1987;9:665–92.
- [103] Morelli G, Malorny B, Muller K, Seiler A, Wang JF, del Valle J, et al. Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. *Mol Microbiol* 1997;25:1047–64.
- [104] Moore PS, Reeves MW, Schwartz B, Gellin BG, Broome CV. Intercontinental spread of an epidemic group A *Neisseria meningitidis* strain. *Lancet* 1989;2:260–3.
- [105] Nicolars P, Norheim G, Garnotel E, Djibo S, Caugant DA. Molecular epidemiology of *Neisseria meningitidis* isolated in the African Meningitis Belt between 1988 and 2003 shows dominance of sequence type 5 (ST-5) and ST-11 complexes. *J Clin Microbiol* 2005;43:5129–35.
- [106] Frasch CE, Preziosi MP, LaForce FM. Development of a group A meningococcal conjugate vaccine, MenAfriVac(TM). *Hum Vaccin Immunother* 2012;8:715–24.
- [107] Novak RT, Kambou JL, Diomande FV, Tarbando TF, Ouedraogo-Traore R, Sangare L, et al. Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. *Lancet Infect Dis* 2012;12:757–64.
- [108] Aguilera JF, Perrocheau A, Meffre C, Hahne S, Group WW. Outbreak of serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. *Emerg Infect Dis* 2002;8:761–7.
- [109] Popovic T, Sacchi CT, Reeves MW, Whitney AM, Mayer LW, Noble CA, et al. *Neisseria meningitidis* serogroup W135 isolates associated with the ET-37 complex. *Emerg Infect Dis* 2000;6:428–9.
- [110] Njanpop-Lafourcade BM, Parent du Chatelet I, Sanou O, Alonso JM, Taha MK. The establishment of *Neisseria meningitidis* serogroup W135 of the clonal complex ET-37/ST-11 as an epidemic clone and the persistence of serogroup A isolates in Burkina Faso. *Microbes Infect* 2005;7:645–9.
- [111] Sacchi CT, Tondella ML, de Lemos AP, Gorla MC, Berto DB, Kumiochi NH, et al. Characterization of epidemic *Neisseria meningitidis* serogroup C strains in several Brazilian states. *J Clin Microbiol* 1994;32:1783–7.
- [112] Wang JF, Caugant DA, Morelli G, Koumare B, Achtman M. Antigenic and epidemiologic properties of the ET-37 complex of *Neisseria meningitidis*. *J Infect Dis* 1993;167:1320–9.
- [113] Halperin SA, Bettinger JA, Greenwood B, Harrison LH, Jelfs J, Ladhani SN, et al. The changing and dynamic epidemiology of meningococcal disease. *Vaccine* 2012;30(Suppl. 2):B26–36.
- [114] Whalen CM, Hockin JC, Ryan A, Ashton F. The changing epidemiology of invasive meningococcal disease in Canada, 1985 through 1992. Emergence of a virulent clone of *Neisseria meningitidis*. *JAMA* 1995;273:390–4.
- [115] Harrison LH, Shutt KA, Schmink SE, Marsh JW, Harcourt BH, Wang X, et al. Population structure and capsular switching of invasive *Neisseria meningitidis* isolates in the pre-meningococcal conjugate vaccine era—United States, 2000–2005. *J Infect Dis* 2010;201:1208–24.
- [116] Jordens JZ, Slack MP. *Haemophilus influenzae*: then and now. *Eur J Clin Microbiol Infect Dis* 1995;14:935–48.
- [117] Gilsdorf JR. Antigenic diversity and gene polymorphisms in *Haemophilus influenzae*. *Infect Immun* 1998;66:5053–9.
- [118] Gilsdorf JR, Marrs CF, Foxman B. *Haemophilus influenzae*: genetic variability and natural selection to identify virulence factors. *Infect Immun* 2004;72:2457–61.
- [119] Jordan IK, Conley AB, Antonov IV, Arthur RA, Cook ED, Cooper GP, et al. Genome sequences for five strains of the emerging pathogen *Haemophilus haemolyticus*. *J Bacteriol* 2011;193:5879–80.
- [120] McCrea KW, Xie J, LaCross N, Patel M, Mukundan D, Murphy TF, et al. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. *J Clin Microbiol* 2008;46:406–16.
- [121] Murphy TF, Brauer AL, Sethi S, Kilian M, Cai X, Lesse AJ. *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. *J Infect Dis* 2007;195:81–9.

- [122] Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, Maiden MC, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 2002;52:1043–7.
- [123] Hedegaard J, Okkels H, Bruun B, Kilian M, Mortensen KK, Nørskov-Lauritsen N. Phylogeny of the genus *Haemophilus* as determined by comparison of partial *infB* sequences. *Microbiology* 2001;147:2599–609.
- [124] Meats E, Feil EJ, Stringer S, Cody AJ, Goldstein R, Kroll JS, et al. Characterization of encapsulated and nonencapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol* 2003;41:1623–36.
- [125] Nørskov-Lauritsen N, Bruun B, Kilian M. Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol* 2005;55:449–56.
- [126] Drancourt M, Signoli M, Dang LV, Bizot B, Roux V, Tzortzis S, et al. *Yersinia pestis* Orientalis in remains of ancient plague patients. *Emerg Infect Dis* 2007;13:332–3.
- [127] Haensch S, Bianucci R, Signoli M, Rajerison M, Schultz M, Kacki S, et al. Distinct clones of *Yersinia pestis* caused the Black Death. *PLoS Pathog* 2010;6:e1001134.
- [128] Raoult D, Aboudharam G, Crubezy E, Larrouy G, Ludes B, Drancourt M. Molecular identification by “suicide PCR” of *Yersinia pestis* as the agent of medieval black death. *Proc Natl Acad Sci USA* 2000;97:12800–3.
- [129] Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 1999;96:14043–8.