The use of increasingly powerful genotyping tools for the characterisation of pathogens has become a standard component of infectious disease surveillance and outbreak investigations. This thematic issue of *Eurosurveillance*, published in two parts, provides a series of review and original research articles that gauge progress in molecular epidemiology strategies and tools, and illustrate their applications in public health. Molecular epidemiology of infectious diseases combines traditional epidemiological methods with analysis of genome polymorphisms of pathogens over time, place and person across human populations and relevant reservoirs, to study host–pathogen interactions and infer hypotheses about host-to-host or source-to-host transmission [1-3]. Based on discriminant genotyping of human pathogens, clonally derived strains can be identified as likely links in a chain of transmission [1-3]. In this two-part issue of *Eurosurveillance*, Goering et al. explain that such biological evidence of clonal linkage complements but does not replace epidemiological evidence of person-to-person contact or common exposure to a potential source [3]. Muellner et al. provide clear examples how prediction about infectious disease outcome and transmission risks can be enhanced through integration of pathogen genetic information and epidemiological modelling to inform public health decisions about food-borne disease prevention [4].

As reviewed by Sabat et al., epidemi source tracing requires timely deployment of high resolution typing methods that index variation of genomic elements with a fast molecular clock [1-5]. For outbreak studies, comparative methods, as opposed to library typing methods, are sufficient, and the higher the power to resolve micro-evolutionary distance, the greater the likelihood to decide between alternative transmission hypotheses generated by observational epidemiology [1-6]. Once standardised to enable a uniform genotype nomenclature across laboratories, thereby providing a library typing system, such discriminatory methods can be further applied to control-oriented surveillance [1-5]. Early outbreak detection is achieved by genotyping prospectively as many consecutive cases in a population as possible to identify clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5]. Examples include clusters of clonally linked isolates [5].

As summarised by Sabat et al., there have been continuous technological improvements for microbial genomic characterisation in the past decade, moving from fingerprinting methods such as pulsed-field gel electrophoresis of bacterial macrorestriction fragments to more robust, portable and biologically informative assays such as bacterial multilocus variable-number tandem repeat analysis (MLVA) and sequencing of single/multiple loci of both bacterial and viral human pathogens [3-5,9-11]. With the decreasing cost and continuing refinement of high-throughput
genome sequencing technologies, we are now witnessing a quantum leap from genotypic epidemiology to genomic epidemiology as whole viral or bacterial genomes become open to scrutiny at population level. As reviewed by Carrico et al., advances in laboratory typing tools have been enabled by parallel progress in the information technology needed to capture genetic data on pathogens, and in quality control, formatting, storage, management and, most importantly, bioinformatics analysis and real-time electronic data sharing through online databases [10].

Among the sequence-based genotyping assays, MLST is widely applied for epidemiological investigations of bacterial and fungal pathogens and is a primary typing method for clonal delineation in pathogens such as Neisseria [12] or Campylobacter [4]. The advantages of MLST are twofold: firstly, it generates reproducible and standardised data that are highly portable (i.e. easily transferrable between different systems) and comparable across laboratories in centralised databases accessible through the Internet. Secondly, the nucleotide substitutions that underlie MLST variation can be interpreted directly in terms of population genetics and evolutionary processes. Because nucleotide polymorphisms evolve slowly in bacteria, MLST is very appropriate to describe the patterns of genetic variation within bacterial species at the global scale. Therefore, one of the major applications of MLST is to decipher bacterial population structure, including clonal diversity, to create a phylogenetic structure of different lineages and to assess the impact of homologous recombination. Recently, this has led to a bold proposal to replace the 70 year-old serotyping nomenclature system for Salmonella strains with MLST [13].

To reduce costs and increase speed, typing based on the sequencing of single highly variable genes was developed for a few pathogens. The most widely used systems are sequencing of the emm gene coding for the M antigen of Streptococcus pyogenes (which can be compared to the results from traditional M serotyping) and the spa gene coding for surface protein A of Staphylococcus aureus [5]. However, single locus typing approaches are limited by events such as homoplasy (evolutionary reversion or convergence) and horizontal gene transfer, as discussed by Sabat et al. [5].

Lindstedt et al. show in this issue how interest in MLVA has grown from the limitations of MLST and other methods to discriminate among isolates of epidemiologically important clones, such as Escherichia coli O157:H7 and Salmonella serovar Typhimurium [11]. MLVA retains the ‘multilocus’ concept of MLST but is based on rapidly evolving loci characterised by the presence of short, tandem repeated sequences. MLVA has proven very useful in surveillance and epidemiology, e.g. for monitoring clonal trends, cluster detection and outbreak investigation [5,11,14]. The high discriminatory power of MLVA for many bacterial groups, combined with its simplicity, makes it an especially useful subtyping tool for so-called monomorphic pathogens [5,11]. In addition, MLVA has a strong potential for inter-laboratory standardisation, and several web-accessible database systems have been developed [5,10-11]. One important drawback is that many MLVA schemes are highly specific for given clones, thus limiting their applicability. Furthermore, for long-term epidemiology or population biology, MLVA markers can be affected by homoplasy, which renders MLVA data less robust than MLST as a library typing system and for phylogenetic purposes. It also remains unclear whether assembly of high throughput sequence data will be reliable enough to determine MLVA alleles, as the repeat arrays pose particular technical challenges for current high throughput sequencing technologies.

From a perspective of medical and public health microbiology and epidemiology, whole genome sequencing (WGS) combines two decisive advantages compared to previous methods: it provides maximal strain discrimination on the one hand, and can be linked to clinically and epidemiologically relevant phenotypes on the other hand. The method is widely seen as the ultimate tool for epidemiological typing of bacteria and other pathogens. It has already proven highly informative to resolve local S. aureus outbreaks [6] as well as elucidate the evolutionary events leading to the emergence and global dissemination of super-pathogen clones with enhanced virulence and multidrug resistance, such as Clostridium difficile ribotype 027 strains [14-15]. Moreover, WGS will provide full genomic characteristics of the infectious isolates, including the set of genes linked to antimicrobial resistance (the resistome) and those linked to virulence of the isolates (the virulome). As discussed by several authors in this issue [3,5,10,12,14], WGS still remains to be fully harnessed conceptually and fine-tuned technologically. This promising technology currently faces three major challenges: speed, data analysis and interpretation, and cost.

As opposed to previous sequence-based typing methods, WGS will change the way we look at pathogen diversity in one fundamental way: without an a priori focus on a subset of loci. As all genetic information will be available, it will allow the discovery of novel, unexpected variation, including polymorphisms that evolve during outbreaks or changes that are selected in vivo during infection. Such pathoadaptive changes can result in increased virulence or novel pathophysiological processes. One example of such a micro-evolutionary change is the emergence during influenza A(H1N)pdm09 epidemic of a quasispecies variant with a haemagglutinin D222G mutation which is associated with modified tissue receptor tropism and severe influenza virus infections, as reported by Rykkvin et al. in this journal [16]. Due to the rapid rate of evolution of viruses and their small genomes, virologists have long been using genome-wide sequencing. The term ‘phylogeny’ designates the study of the interplay of epidemiological and evolutionary patterns, pioneered in
virology [17]. Phylodynamics based on WGS of bacterial populations is emerging as a fertile field of investigation for public health microbiology [5-6,14-15].

As discussed by Jolley and Maiden, WGS sequencing of bacterial pathogens and archiving of the collected data will raise the issue of genomic strain nomenclature [12]. One particularly interesting advantage of MLST in the era of high-throughput sequencing lies in its forward compatibility with future whole genome sequencing, or core genome allotyping, as underlined by Sabat et al. and Jolley and Maiden [5,12]. Several recent tools allow extracting MLST information from high-throughput sequencing data [12,18,19]. The BIGSDB bioinformatics application incorporates MLST databases and provides the possibility to extend the MLST approach to include the full core genome [12]. We anticipate that a WGS-based genotype nomenclature could be developed as a complement to the well-established MLST nomenclature of bacterial clones. As core genome evolution within MLST clones is mainly mutational, the possibility to reconstruct phylogeny based on WGS data should allow a hierarchical classification of WGS types, giving access to different levels of genetic distance resolution depending on the epidemiological questions and length of the study period. This is just one example of the challenges that we face as we enter the exciting era of genomic epidemiology [5,10,12].

Beyond the hurdles in technology and bioinformatics that we still need to overcome, what are the needs for translating advances in genomic epidemiology into public health benefits? Laboratory-based surveillance is pivotal to monitoring infectious disease threats to human health. It relies on aggregating microbiological data that are produced at clinical care level and supplemented by reference laboratory testing. As highlighted by Niesters et al., molecular methods supplant culture-based diagnostic methods, thereby making genomic information relevant to disease surveillance available at the level of the diagnostic laboratory. This technological shift challenges the hierarchical architecture of surveillance networks that relies on samples and culture specimens being referred from the clinics to the reference laboratories and public health institutes [20]. Niesters et al. describe the pilot experience with the TYPENED surveillance network as a molecular data-sharing platform pioneered in the Netherlands by a consortium of clinics, academic institutions and public health virology laboratories [20]. This collaborative approach led to a consensus on how to choose surveillance targets, harmonise sequence-based virological diagnostic assays and share sequence data through a common platform [20].

In addition to stimulating changes in public health systems, the application of high-resolution typing tools such as WGS in outbreak management raises a number of ethical questions, as discussed by Rump et al. in this journal [21]: protection of personal data, informed consent with regard to the investigation of clinical samples, and moral responsibility and legal liability to act upon the evidence to prevent or mitigate disease transmission. As real-time data sharing becomes technically feasible for surveillance and cross-border outbreak investigations, public health organisations will need to develop a policy for the use of these data that balances risks and benefits and defines adequate governance. As part of its mandate to foster collaboration between expert and reference laboratories supporting prevention and control of infectious diseases, the European Centre for Disease Prevention and Control (ECDC) is facilitating interdisciplinary collaboration and assessing public health needs for the integration of microbial genotyping data into surveillance and epidemic preparedness at European level [22]. As announced recently, a European data exchange platform that combines typing data with epidemiological data on a list of priority diseases is being piloted for molecular surveillance of multidrug-resistant *Mycobacterium tuberculosis* and food-borne pathogens [23]. As WGS gradually becomes part of epidemiological studies, ECDC is party to the international expert consultations aimed at building interoperable databases of microbial genomes for future application in public health [24].
References


