

# Overview of molecular typing methods for outbreak detection and epidemiological surveillance

A J Sabat<sup>1</sup>, A Budimir<sup>2</sup>, D Nashev<sup>3</sup>, R Sá-Leão<sup>4</sup>, J M van Dijl<sup>1</sup>, F Laurent<sup>5</sup>, H Grundmann<sup>1</sup>, A W Friedrich (alex.friedrich@umcg.nl)<sup>1</sup>, on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM)<sup>6</sup>

1. Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
2. Department of Clinical Microbiology and Molecular Microbiology, Clinical Hospital Centre Zagreb, Zagreb, Croatia
3. Department of Microbiology, National Center of Infectious and Parasitic Diseases, Sofia, Bulgaria
4. Laboratory of Molecular Microbiology of Human Pathogens, Instituto de Tecnologia Química e Biológica, Oeiras, Portugal
5. Department of Bacteriology, National Reference Centre for Staphylococci, Inserm U81, Hospices Civils de Lyon, University of Lyon, Lyon, France
6. European Society for Clinical Microbiology and Infectious Diseases, Basel, Switzerland

## Citation style for this article:

Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW, on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill.* 2013;18(4):pii=20380. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20380>

Article submitted on 30 June 2012 / published on 24 January 2013

Typing methods for discriminating different bacterial isolates of the same species are essential epidemiological tools in infection prevention and control. Traditional typing systems based on phenotypes, such as serotype, biotype, phage-type, or antibiogram, have been used for many years. However, more recent methods that examine the relatedness of isolates at a molecular level have revolutionised our ability to differentiate among bacterial types and subtypes. Importantly, the development of molecular methods has provided new tools for enhanced surveillance and outbreak detection. This has resulted in better implementation of rational infection control programmes and efficient allocation of resources across Europe. The emergence of benchtop sequencers using next generation sequencing technology makes bacterial whole genome sequencing (WGS) feasible even in small research and clinical laboratories. WGS has already been used for the characterisation of bacterial isolates in several large outbreaks in Europe and, in the near future, is likely to replace currently used typing methodologies due to its ultimate resolution. However, WGS is still too laborious and time-consuming to obtain useful data in routine surveillance. Also, a largely unresolved question is how genome sequences must be examined for epidemiological characterisation. In the coming years, the lessons learnt from currently used molecular methods will allow us to condense the WGS data into epidemiologically useful information. On this basis, we have reviewed current and new molecular typing methods for outbreak detection and epidemiological surveillance of bacterial pathogens in clinical practice, aiming to give an overview of their specific advantages and disadvantages.

## Introduction

Identifying different types of organisms within a species is called typing. Traditional typing systems based on phenotype, such as serotype, biotype, phage-type or antibiogram, have been used for many years. However, the methods that examine the relatedness of isolates at a molecular level have revolutionised our ability to differentiate among bacterial types (or subtypes). The choice of an appropriate molecular typing method (or methods) depends significantly on the problem to solve and the epidemiological context in which the method is going to be used, as well as the time and geographical scale of its use. Importantly, human pathogens of one species can comprise very diverse organisms. Therefore, typing techniques should have excellent typeability to be able to type all the isolates studied [1]. In outbreak investigations, a typing method must have the discriminatory power needed to distinguish all epidemiologically unrelated isolates. Ideally, such a method can discriminate very closely related isolates to reveal person-to-person strain transmission, which is important to develop strategies to prevent further spread. At the same time it must be rapid, inexpensive, highly reproducible, and easy to perform and interpret [1,2]. When typing is applied for continuous surveillance, the respective method must yield results with adequate stability over time to allow implementation of efficient infection control measures. Moreover, a typing method that is going to be used in international networks should produce data that are portable (i.e. easily transferrable between different systems) and that can be easily accessed via an open source web-based database, or a client-server database connected via the Internet. Additionally, a typing method used for surveillance should rely on an internationally standardised nomenclature, and it should be applicable for a broad range of bacterial species. There should also

be procedures in place to check and validate, by using quantifiable internal and external controls, that the typing data are of high quality. A clear advantage for a typing approach is the availability of software that: (i) enables automated quality control of raw typing data, (ii) allows pattern/type assignment, (iii) implements an algorithm for clustering of isolates based on the obtained data, (iv) provides assistance in the detection of outbreaks of infections, and (v) facilitates data management and storage. To date, many different molecular methods for epidemiological characterisation of bacterial isolates have been developed. However, none of them is optimal for all forms of investigation. Thus, a thorough understanding of the advantages and limitations of the available typing methods is of crucial importance for selecting the appropriate approaches to unambiguously define outbreak strains.

Here, we present an overview of the typing methods that are currently used in bacterial disease outbreak investigations and active surveillance networks, and we specify their advantages and disadvantages. Importantly, we focus on those methods that have the strongest impact on public health, or for which there is a growing interest in relation to clinical use.

### PubMed database searches

To investigate the impact of typing methods in public health, we first queried the PubMed database using a combination of specific keywords to retrieve the relevant articles without any constraints on the time of publication. Furthermore, in order to reveal a growing interest in particular typing methods, we subsequently restrictively searched PubMed for articles published between January 2010 and the present day (as of 1 December 2012). We considered a method as a method of growing interest when the number of articles published between January 2010 and the present day was higher than the number of articles published before 2010. Specifically, an electronic search was conducted using the following combinations of keywords: PFGE [AND] typing; AFLP [AND] typing; RAPD [AND] typing; DiversiLab [AND] typing; VNTR [AND] typing; *emm* [OR] *flab* [AND] typing; *spa* [AND] typing; MLST [AND] typing; whole [AND] genome [AND] sequencing [AND] typing; microarrays [OR] microarray [AND] typing; optical [OR] whole [AND] genome [AND] mapping [AND] typing. Also, to identify the impact of particular typing methods on outbreak investigations currently conducted, we searched the PubMed database with a restriction to articles published between January 2011 and the present day, using the following combinations of specific keywords: PFGE [AND] outbreak; AFLP [AND] typing; RAPD [AND] typing; DiversiLab [AND] outbreak; VNTR [AND] outbreak; *emm* [OR] *flab* [AND] outbreak; *spa* [AND] typing [AND] outbreak; MLST [AND] outbreak; whole [AND] genome [AND] sequencing [AND] outbreak; microarrays [AND] outbreak; optical [OR] whole [AND] genome [AND] mapping [AND] outbreak. The results of these literature searches have been included

in the following sections of this review that address the respective typing methods.

### Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) has been considered as the 'gold standard' among molecular typing methods for a variety of clinically important bacteria. When 'PFGE AND typing' were used as search terms, over 2,700 publications were retrieved in PubMed, which underscores the major influence and importance of this method in the field. For most bacterial species, the technique was adopted as an epidemiological tool in the 1990s [3-6]. Today, it is still the most frequently used approach to characterise bacterial isolates in outbreaks [7,8] as revealed by a PubMed database search with a restriction to articles published between January 2011 and the present day. In total, 183 hits were obtained for the terms 'PFGE AND outbreak', while searches for all other methods in combination with the term 'outbreak' invariably resulted in less than 100 hits. For many years, PFGE has been a primary typing tool to analyse centre-to-centre transmission events, and it has been used successfully in large-scale epidemiological investigations [9]. The success of PFGE results from its excellent discriminatory power and high epidemiological concordance. Moreover, it is a relatively inexpensive approach with excellent typeability and intra-laboratory reproducibility. In the past decade, protocols for PFGE have been standardised and inter-laboratory comparison has been undertaken through several initiatives, such as PulseNet [10] or Harmony [11]. It has also been possible to establish international fingerprinting databases, which allowed fast detection of emerging clones and monitoring of the spread of pathogenic bacterial strains through different regions or countries. To perform PFGE, a highly purified genomic DNA sample is cleaved with a restriction endonuclease that recognises infrequently occurring restriction sites in the genome of the respective bacterial species. The resulting restriction fragments, which are mostly large, can be separated on an agarose gel by 'pulsed-field' electrophoresis in which the orientation of the electric field across the gel is changed periodically. The separated DNA fragments can be visualised on the gel as bands, which form a particular pattern on the gel, the PFGE pattern. For most bacteria PFGE can resolve DNA fragments with sizes ranging from about 30 kb to over 1 Mb [12]. Large restriction fragments are thus separated in a size-dependent manner and the method yields relatively few bands on the gel, which makes analysis of the results easier. A clear advantage of the PFGE method is that it addresses a large portion of an investigated genome (>90%). Accordingly, insertions or deletions of mobile genetic elements as well as large recombination events within genomic DNA will result in changes in the PFGE patterns. Usually, plasmid DNA does not interfere with the macrorestriction profiles of the chromosomal DNA, which is responsible for the particular PFGE pattern, as the fragments generated by restriction of plasmid DNA are too small to affect the profile. However,

in some bacteria, differences in the carriage of large plasmids (over 50 kb) have been observed as single-band differences between the respective PFGE profiles [12]. Unfortunately, although widely used, PFGE suffers from several limitations. The method is technically demanding, labour-intensive and time-consuming, and it may lack the resolution power to distinguish bands of nearly identical size (i.e. fragments differing from each other in size by less than 5%). Moreover, the analysis of PFGE results is prone to some subjectivity and the continuous quality control and portability of data are limited compared to sequence-based methods.

### **Amplified fragment length polymorphism**

In the amplified fragment length polymorphism (AFLP) method, genomic DNA is cut with two restriction enzymes, and double-stranded adaptors are specifically ligated to one of the sticky ends of the restriction fragments [13]. Subsequently, the restriction fragments ending with the adaptor are selectively amplified by polymerase chain reaction (PCR) using primers complementary to the adaptor sequence, the restriction site sequence and a number of additional nucleotides (usually 1–3 nucleotides) from the end of the unknown DNA template. At the start of the amplification, highly stringent conditions are used to ensure efficient binding of primers to fully complementary nucleotide sequences of the template. AFLP allows the specific co-amplification of high numbers (typically between 50 and 100) of restriction fragments and is often carried out with fluorescent dye-labeled PCR primers. This allows to detect the fragments once they have been separated by size on an automated DNA sequencer. A subsequent computer-assisted comparison of high-resolution banding patterns generated during the AFLP analysis enables the determination of genetic relatedness among studied bacterial isolates [14]. AFLP has been described as being at least as discriminatory as PFGE [15]. In addition, AFLP is a reproducible approach and like other DNA banding pattern-based methods it can be automated [16] and results are portable. The major limitations of AFLP include the fact that it is labour-intensive (a typical analysis takes about three days), and the kits for extraction of the total DNA, enzymes, fluorescence detection systems and adaptors are expensive.

### **Random amplification of polymorphic DNA and arbitrarily primed polymerase chain reaction**

Random amplification of polymorphic DNA (RAPD) is based on the parallel amplification of a set of fragments by using short arbitrary sequences as primers (usually 10 bases) that target several unspecified genomic sequences. Amplification is conducted at a low, non-stringent annealing temperature, which allows the hybridisation of multiple mismatched sequences. When the distance between two primer binding sites on both DNA strands is within the range of 0.1–3 kb, an amplicon can be generated that covers the sequence between these two binding sites. Importantly, the number and the positions of primer binding sites are

unique to a particular bacterial strain. RAPD amplicons can be analysed by agarose gel electrophoresis or DNA sequencing depending on the labeling of primers with appropriate fluorescent dyes. Although, less discriminatory than PFGE, RAPD has been widely used for the typing of bacterial isolates in cases of outbreaks [17,18], because it is simple, inexpensive, rapid and easy in use. The main drawback of the RAPD method is its low intra-laboratory reproducibility since very low annealing temperatures are used. Moreover, RAPD lacks inter-laboratory reproducibility since it is sensitive to subtle differences in reagents, protocols, and machines.

Arbitrarily primed PCR (AP-PCR) is a variant of the original RAPD method, and it is therefore often referred to as RAPD [19]. The differences between the AP-PCR and RAPD protocols involve several technical details. In AP-PCR: (i) the amplification is conducted in three parts, each with its own stringency and concentration of components, (ii) high primer concentrations are used in the first PCR cycles, and (iii) primers of variable length and often designed for other purposes are used. Consequently, the advantages and limitations of AP-PCR are identical to those of RAPD, as pointed out above.

### **Repetitive-element polymerase chain reaction**

Repetitive-element PCR (rep-PCR) is based on genomic fingerprint patterns to classify bacterial isolates. The rep-PCR method uses primers that hybridise to non-coding intergenic repetitive sequences scattered across the genome. DNA between adjacent repetitive elements is amplified using PCR and multiple amplicons can be produced, depending on the distribution of the repeat elements across the genome. The sizes of these amplicons are then electrophoretically characterised, and the banding patterns are compared to determine the genetic relatedness between the analysed bacterial isolates. Multiple families of repeat sequences have been used successfully for rep-PCR typing, such as the ‘enterobacterial repetitive intergenic consensus’ (ERIC), ‘the repetitive extragenic palindromic’ (REP), and the ‘BOX’ sequences [20]. As this typing approach is based on PCR amplification and subsequent DNA electrophoresis, the results of rep-PCR can be obtained in a relatively short period of time. This is also the reason why this approach is very cheap. For many bacterial organisms rep-PCR can be highly discriminatory [21,22]. The main limitation of rep-PCR combined with electrophoresis using traditional agarose gels is that it lacks sufficient reproducibility, which may result from variability in reagents and gel electrophoresis systems.

The DiversiLab system (bioMérieux, Marcy l’Etoile, France) is a semiautomated method using the rep-PCR approach. We mention it here, because it is used in local infection control settings by a number of hospitals worldwide. In this case, commercial PCR kits have been developed for a series of clinically important

microorganisms [23]. After PCR, amplified genomic DNA regions between repetitive elements are separated by high-resolution chip-based microfluidic capillary electrophoresis. The microfluidic capillary electrophoresis has been utilised by the DiversiLab system to substantially increase resolution and reproducibility of the rep-PCR approach in comparison to traditional gel electrophoresis. The resulting data are automatically collected, normalised and analysed by the DiversiLab software. A number of studies have evaluated the usefulness of DiversiLab by comparing its performance with current standard typing methods using well-characterised collections of outbreak-related and epidemiologically unrelated bacterial isolates [24-26]. These studies have shown that the DiversiLab system is simple, easy to perform, rapid, reproducible, endowed with full typeability and applicable to a wide range of microorganisms. The authors concluded that for most bacterial species, in case of a suspected outbreak in hospital settings, DiversiLab is useful especially in first-line outbreak detection. In particular, Fluit and colleagues [25] have shown that DiversiLab is a useful tool for identification of hospital outbreaks of *Acinetobacter* spp., *Stenotrophomonas maltophilia*, *Enterobacter cloacae*, *Klebsiella* spp., and *Escherichia coli*, but that it is inadequate for *Pseudomonas aeruginosa*, *Enterococcus faecium*, and methicillin-resistant *Staphylococcus aureus* (MRSA). The view that DiversiLab can be insufficiently discriminative for typing some bacterial species, including MRSA, in outbreak settings was confirmed by Babouee et al. [27]. The results obtained by Overdeest and colleagues [26], who evaluated the performance of DiversiLab, were also in line with the findings reported by Fluit et al. [25], except for the conclusions regarding *P. aeruginosa*. Deplano and colleagues [24] have demonstrated excellent epidemiological concordance of the results produced by DiversiLab by correctly linking all outbreak-related isolates of vancomycin-resistant *E. faecium* (VREF), *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *P. aeruginosa*. However, they also recommended that for *E. coli* isolates with the same DiversiLab type, the results should be confirmed by testing additional markers [24]. The total cost of all consumables and reagents for DiversiLab is comparable to that of PFGE, amounting in euros (EUR) to about EUR 20 per isolate. By checking the PubMed database using 'DiversiLab AND typing' as the search term, 63 publications were retrieved of which 48 were dated after the end of 2009. This indicates a growing interest in the use of DiversiLab as a typing tool. However, as the inter-laboratory reproducibility of rep-PCR approaches is generally limited, large-scale intra- and inter-laboratory reproducibility studies should be carefully performed to further evaluate the usefulness of the DiversiLab system for regional and eventually national surveillance of bacterial genotypes. Moreover, the DiversiLab database is housed on a manufacturer server, which prevents some potential users from using this typing system because of concerns with data security issues.

## Variable-number tandem repeat (VNTR) typing

Bacterial genomes possess many regions with nucleotide repeats in coding and non-coding DNA sequences. When these repeats are directly adjacent to each other and their number at the same locus varies between isolates, the respective genomic regions are called variable-number tandem repeat (VNTR) loci. The repeats at the same locus can be identical or their nucleotide sequences can differ slightly. Multilocus VNTR analysis (MLVA) is a method which determines the number of tandem repeat sequences at different loci in a bacterial genome. In a most simple MLVA assay, a number of well-selected VNTR loci are amplified by multiplex PCR and an analysis of the amplicons is conducted on standard agarose gels [28]. An advantage of this simple but also cheap, fast and easy to use assay is that the whole procedure can be performed in laboratories without sophisticated electrophoresis equipment. When MLVA does not enable a convenient and unambiguous calculation of the individual numbers of repeats per locus, some investigators call it multiple-locus VNTR fingerprinting (MLVF) [21,29]. A drawback of MLVF is that the resulting data cannot be compared directly between different laboratories. This is due to the fact that the generated amplicons are monitored as banding patterns by conventional electrophoresis on low-resolution agarose gels. Such analyses do not reveal the exact numbers of repeats in the obtained amplicons and it is also impossible to determine which band in a pattern corresponds to which PCR target. A better separation of the amplified DNA fragments by size during electrophoresis has been achieved by replacement of standard agarose gels with a microfluidic chip-based analysis on a fully integrated miniaturised instrument. In 2005, Francois and colleagues [30] reported on the use of automated microfluidic electrophoresis with the Agilent 2100 bioanalyzer 'lab-on-a-chip' for the VNTR typing of *S. aureus* isolates. Since then, there have been a growing number of studies that have shown the clear advantage of microfluidic chips over the standard agarose gels for the MLVA/MLVF typing in terms of electrophoretic separation resolution, reproducibility, rapidity and automated data analysis [31,32].

For inter-laboratory comparison, the exact number of repeat units in each MLVA locus must be determined. From the size of a particular PCR product and the known length of a single repeat and the flanking consensus regions to which primers were designed, the number of repeated units at each locus can be calculated. The use of capillary electrophoresis on an automatic DNA sequencer and the labeling of primers with different fluorescently coloured dyes allows MLVA amplicons to be analysed in one run and still be typed individually [33,34]. The different fluorophore molecules incorporated in the amplicons absorb the laser energy and release light of different wavelengths, which are then identified by the detector in the DNA sequencer. Using computer software, all loci are distinctly recognised

on electropherograms according to their colours, and based on their amplicon sizes, the repeat number per MLVA locus is calculated automatically. Moreover, the determination of amplicon sizes using a DNA sequencer is conducted much more precisely than when agarose gels or microfluidic chips are used. Once the number of repeats in a set of VNTR loci (alleles) for a bacterial isolate is assessed, an ordered string of allele numbers corresponding to the number of repeat units at each MLVA locus results in an allelic profile (e.g. 7-12-3-3-22-11-6-1), which can be easily compared to reference databases via the Internet.

The intrinsic limitation of MLVA is that it is not a universal method, meaning that primers need to be designed specifically for each pathogenic species targeted. This is the major reason why it cannot replace PFGE in epidemiological investigations in general. Furthermore, MLVA is not 100% reproducible unless the allele amplicons are sequenced and the users have agreed on where the VNTR begins and ends for each locus. For improved reproducibility of MLVA, single PCR amplifications of VNTR loci instead of multiplex reactions can be conducted. However, this approach increases the assay time and its costs. Separation by size of amplicons is not reproducible when using different sequencers, polymers, or fluorescent labels. The size difference in a VNTR locus may not always reflect the real number of tandem repeats, because insertions, deletions or duplications in the amplified region can also give rise to the same size difference. Therefore, sequencing of the amplicons is necessary in this case. Importantly, MLVA has not yet been fully developed and properly validated for use in surveillance networks dedicated to clinically relevant organisms as is underscored by the fact that multiple protocols have been published that still remain to be carefully validated.

An alternative strategy for epidemiological typing is the measurement of variations in the VNTR regions by DNA sequencing. Methods relying on sequence variations in multiple VNTR regions have been developed for the subtyping of *Mycobacterium avium* subsp. *paratuberculosis* [35], *Vibrio cholerae* [36], and *Legionella pneumophila* [37] isolates.

When 'VNTR AND typing' were used as a search term in PubMed, about 1,000 publications were retrieved from PubMed, showing that VNTR-based typing approaches are of major importance in the field.

### Single locus sequence typing

Single locus sequence typing (SLST) is used to determine the relationships among bacterial isolates based on the comparison of sequence variations in a single target gene. The terminology SLST has been borrowed from the better known approach called multilocus sequence typing (MLST) (see below) in which several genes are characterised by DNA sequencing to determine genetic relatedness among the isolates.

Typing based on the M-protein found on the surface of group A *Streptococcus* (GAS) has been the most widely used method for distinguishing GAS isolates [38]. The M-protein, encoded by the *emm* gene, is the major virulence and immunological determinant of this human-specific pathogen. In recent years, the classic M-protein serological typing was largely replaced by sequencing of the hypervariable region located at the 5' end of the *emm* gene [39]. The *emm*-typing method has become the gold-standard of GAS molecular typing for surveillance and epidemiological purposes, and more than 200 *emm* types have been described so far. Nevertheless, in order to fully discriminate GAS clones, *emm*-typing should be complemented with other typing methods, like PFGE or MLST [40,41].

Nucleotide sequencing of the short variable region (SVR) of the flagellin B gene (*flaB*) provides adequate information for the study of *Campylobacter* epidemiology. Although PFGE remains the most discriminatory typing method for *Campylobacter*, a study conducted by Mellmann and colleagues [42] showed that sequencing of the SVR region of *flaB* is a rapid, reproducible, discriminatory and stable screening tool. It was also found that *flaB* sequence-typing is useful in combination with other typing methods such as MLST to differentiate closely related or outbreak isolates [43].

When 'emm OR *flab* AND typing' were used as a search term in PubMed, 238 hits were retrieved, which shows the importance of this method for the typing of GAS and *Campylobacter* isolates.

### *Staphylococcus aureus* protein A gene-typing

The most widely used method of the SLST group is called *S. aureus* protein A gene (*spa*)-typing, because it involves the sequencing of the polymorphic X region of the protein A gene of *S. aureus*. Molecular typing of *S. aureus* isolates on the basis of the protein A gene polymorphism was the first bacterial typing method based on repeat sequence analysis [44]. The high degree of genetic diversity in the VNTR region of the *spa* gene results not only from a variable number of short repeats (24 bp), but also from various point mutations. In the *spa* sequence typing method, each identified repeat is associated to a code and a *spa*-type is deduced from the order of specific repeats. Although *spa*-typing has a lower discriminatory ability than PFGE [45,46], its cost-effectiveness, ease of use, speed, excellent reproducibility, appropriate *in vivo* and *in vitro* stability, standardised international nomenclature, high-throughput by using the StaphType software, and full portability of data via the Ridom database (<http://spaserver.ridom.de>) makes this method the currently most useful instrument for characterising *S. aureus* isolates at the local, national and international levels [47-52]. Importantly, this approach ensures strict criteria for internal and external quality assurance of data submitted to the database that is curated by SeqNet.org [50,53]. Furthermore, the implementation

of the based upon repeat patterns (BURP) algorithm to the StaphType software has greatly facilitated the assignment of *spa*-types into clonal complexes and singletons. Nevertheless, *spa*-typing has also certain disadvantages. The major drawback of this method based on single-locus typing is that it can misclassify particular types due to recombination and/or homoplasy. When ‘*spa* AND typing’ were used as a search term in PubMed, 548 hits were retrieved, which highlights the importance of this method for the typing of *S. aureus* isolates. Moreover, 341 of the respective publications were dated after the end of 2009, showing that *spa*-typing is gaining an increasing influence.

### Multilocus sequence typing

In order to overcome the lack or poor portability of traditional and older molecular typing approaches, the MLST method has been invented. MLST is based on the principles of phenotypic multilocus enzyme electrophoresis (MLEE) [54], which relies on the differences in electrophoretic mobility of different enzymes present in a bacterium. The first MLST scheme was developed for *Neisseria meningitidis* in 1998 [55]. Shortly thereafter, the method was extended to other bacterial species and, over time, it has become a very popular tool for global epidemiological studies, and for studies on the molecular evolution of pathogens [56-66]. Accordingly, a PubMed search with the term ‘MLST AND typing’ yielded 1,485 hits. In MLST, internal sequences (of approximately 450–500 bp) of mostly seven housekeeping genes are amplified by PCR and sequenced. For each locus, unique sequences (alleles) are assigned arbitrary numbers and, based on the combination of identified alleles (i.e. the ‘allelic profile’), the sequence type (ST) is determined. The number of nucleotide differences between alleles is not considered. The great advantage of MLST is that all data produced by this method are unambiguous due to an internationally standardised nomenclature, and highly reproducible. Moreover, the allele sequences and ST profiles are available in large central databases (<http://pubmlst.org> and [www.mlst.net](http://www.mlst.net)) that can be queried via the Internet. These databases also provide on-line software (eBURST) for determination of the genetic relatedness between bacterial strains within a species as well as MLST-maps to track the isolates of each ST that have been recovered from each country plus the details of these isolates. The great disadvantage of MLST is its high cost. The total costs of all consumables and reagents for MLST greatly depend on the number of loci investigated and the country in which this typing procedure is conducted. We estimate that in Member States of the European Union, the total costs of an MLST analysis based on seven loci amount to about EUR 50 per isolate. In contrast, the total costs of MLVF performed with an Agilent BioAnalyzer, MLVA with a DNA sequencer, or SLST merely amount to about EUR 2, EUR 8 and EUR 8 per isolate, respectively [32]. Moreover, MLST is labour-intensive, time-consuming and for some pathogens insufficiently discriminating for routine use in outbreak investigations and local

surveillance. To increase the discriminatory power of the ‘classical’ MLST schemes based on seven housekeeping genes, the sequencing results for particular antigen-encoding genes can be included in the analysis. This is exemplified, by the two-locus sequence typing (*Neisseria gonorrhoeae* multi-antigen sequence typing, NG-MAST) approach developed for *N. gonorrhoeae*, which includes two of the most variable gonococcal genes, namely *por* and *tbpB* [67]. Another example is the MLST approach developed for *Salmonella enterica* in which two housekeeping genes, *gyrB* and *atpD*, in combination with the flagellin genes *fliC* and *fliB* were applied [68]. Moreover, attempts have been undertaken to develop MLST schemes that are entirely based on virulence genes. Such approaches, termed multi-virulence-locus sequence typing (MVLST), have been applied for the subtyping of pathogens like *Listeria monocytogenes*, *V. cholerae*, *S. enterica* and *S. aureus* [69-72]. Altogether, the currently available data suggest that MVLST is endowed with a higher discriminatory power than that of the ‘classical’ MLST. However, for most of the MVLST approaches, additional research is needed. This should involve different and larger sets of isolates, and the results should also be correlated with conventional epidemiological data in order to validate the applicability of MVLST for epidemiological typing.

### Comparative genomic hybridisation

A DNA microarray used for typing studies is a collection of DNA probes attached in an ordered fashion to a solid surface. These probes can be used to detect the presence of complementary nucleotide sequences in particular bacterial isolates. Thus, microarrays represent facile tools for detecting genes that serve as markers for specific bacterial strains, or to detect allelic variants of a gene that is present in all strains of a particular species. The probes on the array may be PCR amplicons (> 200 bp) or oligonucleotides (up to 70 mers). Depending on the number of probes placed on a solid surface, we can distinguish low-density (hundreds of probes) and high-density (hundreds of thousands of probes) DNA microarrays. In the usual approach, total DNA is extracted from a pathogen of interest. This target DNA is then labeled, either chemically or by an enzymatic reaction, and hybridised to a DNA microarray. Unbound target DNA is removed during subsequent washing steps of different stringency, and the signal from a successful hybridisation event between the labeled target DNA and an immobilised probe is measured automatically by a scanner. The data produced by a microarray assay are then analysed using dedicated software to assess the bacterial diversity. The results retrieved from array technology are variable and depend on the customised array. DNA microarrays appear to be very well suited for bacterial typing as is underscored by the 506 PubMed hits with the search terms ‘microarrays OR microarray AND typing’. Microarrays are currently widely used to analyse genomic mutations, such as single-nucleotide polymorphisms (SNPs). In addition, microarray technology is an efficient tool for the

detection of extra-genomic elements [73,74]. Through microarray-based gene content analyses, pathogens can be simultaneously genotyped and profiled to determine their antimicrobial resistance and virulence potential. Importantly, such a high-density whole genome microarray approach comprises probes allowing for the detection of the open reading frame (ORF) content of one or many genomes. Comparative genomics by using whole genome microarrays has revealed that 10 major *S. aureus* lineages are responsible for the majority of infections in humans [75]. The application of very recently developed microarrays (Sam-62) based on 62 *S. aureus* whole genome sequencing (WGS) projects and 153 plasmid sequences has shown that MRSA transmission events unrecognised by other approaches can be identified using microarray profiling, which is capable of distinguishing between extremely similar but non-identical sequences [73]. Also, a high-density Affymetrix DNA microarray platform based on all ORFs identified on 31 chromosomes and 46 plasmids from a diverse set of *E. coli* and *Shigella* isolates has been applied to quickly determine the presence or absence of genes in very recently emerged *E. coli* O104:H4 and related isolates [76]. This genome-scale genotyping has thus revealed a clear discrimination between clinically, temporally, and geographically distinct O104:H4 isolates. The authors have therefore concluded [76] that the whole genome microarray approach is a useful alternative for WGS to save time, effort and expenses, and it can be used in real-time outbreak investigations. However, the application of high-density microarrays for bacterial typing in routine laboratories is currently hindered by the high costs of materials and the specialised equipment needed for the tests. Alere Technologies has therefore developed a rapid and economic microarray assay for diagnostic testing and epidemiological investigations. The assay was miniaturised to a microtitre strip format (ArrayStrips) allowing simultaneous testing of eight to up to 96 samples. The Alere StaphyType DNA microarray for *S. aureus* covers 334 target sequences, including approximately 170 distinct genes and their allelic variants [77]. Ninety six arrays are scanned on the reader and the affiliation of *S. aureus* isolates to particular genetic lineages is done automatically by software based on hybridisation profiles. With the ArrayStrips, the ArrayTube Platform as a single test format is also available for a number of bacterial species. Interestingly, the total cost of an Alere microarray test per bacterial isolate is comparable to that of PFGE (about EUR 20–30) and much lower than that of MLST (EUR 50). The whole typing procedure for 96 isolates can be conducted within two working days. Recently, Alere Technologies has also developed genotyping DNA microarray kits for other bacterial species, such as *E. coli*, *P. aeruginosa*, *L. pneumophila*, and *Chlamydia trachomatis*. Altogether, the available data show that microarray-based technologies are highly accurate. However, the reproducibility of microarray data within and between different laboratories needs to be established prior to the broad application of this technology. In particular, if SNPs are

the target for typing of highly clonal species, then DNA microarray analysis is probably not the best method to apply. Moreover, arrays have the major disadvantage that they do not allow the identification of sequences which are not included in the array.

Classical serotyping involves a few days to achieve final conclusive results. It requires a major set of costly antisera, is expensive and tedious so that its use is usually restricted to only a few reference laboratories. These technical difficulties can be overcome with molecular serotyping methods. Accordingly, Alere Technologies has developed fast DNA Serotyping assays based on oligonucleotide microarrays for *C. trachomatis*, *E. coli* and *S. enterica* [78,79]. The microarray serogenotyping assay for *C. trachomatis* includes a set of oligonucleotide probes designed to exploit multiple discriminatory sites located in variable domains 1, 2 and 4 of the *ompA* gene encoding the major outer membrane protein A. In case of *E. coli* and *S. enterica*, separate approaches have been developed, but in both these assays the genes encoding the O and H antigens have been selected as target sequences. After multiplex amplification of the selected DNA target sequences using biotinylated primers, the samples are hybridised to the microarray probes under highly stringent conditions. The resulting signals yield genotype (serovar)-specific hybridisation profiles.

### Optical mapping

Optical maps from single genomic DNA molecules were first described for a pathogenic bacterium in the year 2001 [80]. They were constructed for *E. coli* O157:H7 to facilitate genome assembly by an accurate alignment of contigs generated from the large number of short sequencing reads and to validate the sequence data. Optical mapping, also called whole genome mapping, is now a proven approach to search for diversity among bacterial isolates.

Moreover, optical mapping can be coupled with next generation sequencing (NGS) technologies to effectively and accurately close the gaps between sequence scaffolds in *de novo* genome sequencing projects. The system creates ordered, genome-wide, high-resolution restriction maps using randomly selected individual DNA molecules [81]. High molecular weight DNA is obtained from gently lysed cells embedded in low-melting-point agarose. The purified DNA is subsequently stretched on a microfluidic device. Following digestion with a selected restriction endonuclease, the resulting molecule fragments remain attached to the surface of the microfluidic device in the same order as they appear in the genome. The genomic DNA is then stained with an intercalating fluorescent dye and visualised by fluorescence microscopy. The lengths of the restriction fragments are measured by fluorescence intensity. Finally, using specialised software, the consensus genomic optical map is assembled by overlapping multiple single molecule maps. Whole chromosome optical maps can be created for a few organisms within

two days. Due to a very high accuracy and resolution potential, optical mapping has been used successfully in retrospective outbreak investigations to examine the genetic relatedness among isolates of several bacterial species [82-84]. Mellmann and colleagues [85] created for the first time whole chromosome optical maps in real-time outbreak investigations for the *E. coli* isolates recovered from patients in hospitals located in four different German cities during the 2011 outbreak of *E. coli* O104:H4. Based on these studies, it can be concluded that optical mapping is a very powerful tool to assess the genetic relationships among bacterial isolates. However, the use of this technique is currently limited by the high costs of the experiments and the specialised equipment needed.

### Whole genome sequencing

NGS has transformed genetic investigations by providing a cost-effective way to discover genome-wide variations. These NGS technologies are also known as 'second generation sequencing', or 'high-throughput sequencing'. The terms next generation or second generation sequencing are used to distinguish these approaches from the first generation sequencing approaches based on the Sanger method. The clear advantage of NGS over traditional Sanger sequencing is the ability to generate millions of reads (approximately 35–700 bp in length) in single runs at comparatively low costs. To construct the complete nucleotide sequence of a genome, multiple short sequence reads must be assembled based on overlapping regions (*de novo* assembly), or comparisons with previously sequenced 'reference' genomes (resequencing). WGS is becoming a powerful and highly attractive tool for epidemiological investigations [85-88] and it is highly likely that in the near future WGS technology for routine clinical use will permit accurate identification and characterisation of bacterial isolates. However, the key challenge will not be to produce the sequence data, but to rapidly compute and interpret the relevant information from large data sets. Ideally, this information should include and therefore enable a direct comparison to the results obtained by conventional typing methods (e.g. PFGE, MLST), and it should be stored in globally accessible databases. However, the reads produced by the NGS technologies are relatively short, which can make the *de novo* genome assembly a challenging enterprise. Accordingly, the term 'whole genome sequence' refers often to only approximately 90% of the entire genome. The gaps between assembled regions (contigs) are mainly caused by the presence of dispersed or tandemly arrayed repeats.

As current NGS sequencing platforms do not resolve such VNTRs very well, it is often difficult or even impossible to extract useful information on repeats in the MLVA loci from the available genome sequences. Also, for an *in silico* restriction digest to simulate PFGE, there is a need to close completely the gaps between the contigs to obtain one long, contiguous sequence. Therefore, PFGE profiles cannot be predicted without

closing the genome sequences, and on top of this it is necessary to know how different restriction sites used for PFGE are methylated in an organism of interest. To improve *de novo* genome assembly, the introduction of new platforms that generate much longer reads is needed. Recently, a 'third-generation sequencer' (PacBio) was launched by Pacific Biosciences, which generates very long reads with average lengths of 2–3 kb, and reads of more than 7 kb are not uncommon with this system. Furthermore, approximately 100 kb reads are generated by nanopore sequencing technologies as developed by Oxford Nanopore. The main limitations of these third-generation sequencing approaches are their very high costs and low accuracy (approximately 15% error rate). However, further improvements are promised by Pacific Biosystems and Oxford Nanopore to generate long sequence reads with much higher accuracy [89].

The costs of bacterial WGS by NGS continue to decline. Currently, a price level has been reached that comes close to the price of an MLST analysis carried out by traditional Sanger sequencing reactions. Thus, the sequencing cost in United States (US) dollars (USD) of a bacterial genome using NGS can be as little as USD 100–150 per isolate (which amounts to EUR 75–110), including sample preparation, library quality control (quantification and size assessment), and sequencing [90,91]. Not surprisingly, there is an increasing interest in the replacement of PCR/Sanger sequencing with high-throughput deep sequencing technologies, such as 454-pyrosequencing, Illumina and the Ion Torrent system yielding large numbers of short and high-quality reads.

Desktop model sequencers are within the financial reach of many, if not all, reference laboratories. However, the procedure is still too slow, and the genome assembly too complicated for implementation in routine surveillance, as NGS requires heavy computer resources and the help of well-trained bioinformaticians. On the other hand, Windows-based software (e.g. Bionumerics and Lasergene) that does not require deep insights into bioinformatics for assembling the sequenced genomes and query them against reference genomes or other sequences is just around the corner. An important prerequisite for the effective application of WGS technologies in the typing of microorganisms is the availability of novel web-accessible bioinformatics platforms for rapid data processing and analysis. Moreover, these bioinformatics tools should be simple enough for use in clinical settings. This is highly feasible as exemplified by the convenient web-based method for MLST of 66 bacterial species that was developed by Larsen et al. [92]. This method utilises short sequence reads or reassembled genomes for identifying MLST sequence types, and it is publicly available at [www.cbs.dtu.dk/services/MLST](http://www.cbs.dtu.dk/services/MLST).

The great advantage of MLST based on seven house-keeping genes is that this method is fully standardised



for numerous bacterial species. However, a very significant amount of genomic information, including DNA sequence and gene content diversity, exists outside of the genes targeted by traditional MLST. Therefore, to be more effective in the characterisation of outbreak isolates and to strengthen the surveillance systems for particular pathogens, higher resolution methods which utilise WGS are urgently needed. This view is critically underscored by the outbreak of a multidrug-resistant enterohaemorrhagic *E. coli* (EHEC) O104:H4 infection causing a number of haemolytic uraemic syndrome (HUS), which occurred in Germany in the period between May and June 2011 [85,93]. This outbreak resulted in the death of 46 people and more than 4,000 diseased patients [94]. Before the outbreak in 2011, only one case of HUS associated with *E. coli* O104:H4, which took place in 2001, had been reported in Germany [85,95]. The traditional MLST typing based on sequence determination of seven housekeeping genes revealed that both the historical isolate recovered in 2001 and an isolate originating from a HUS patient during the outbreak in 2011 had the same MLST type 678. This indicated that both isolates were closely related. However, in this case, MLST was not able to reveal major differences between the outbreak isolate and the earlier isolate as became clearly evident upon their characterisation by NGS. Strikingly, the WGS data revealed that the isolate originating from the 2011 outbreak differed substantially from the 2001 isolate in chromosomal and plasmid content [85]. An independent study by Hao and colleagues [96] confirmed these results as the analysis of *E. coli* O104:H4 ST678 isolates (one of them was epidemiologically linked to the 2011 outbreak) showed that traditional MLST cannot accurately resolve relationships among genetically related isolates that differ in their pathogenic potentials. Using the WGS data they found in 167 genes an evidence of homologous recombination between distantly related *E. coli* isolates, including the 2011 outbreak isolate [96].

We are convinced that in the near future WGS will become a highly powerful tool for outbreak investigations and surveillance schemes in routine clinical practice. However, this will require standard operating procedures for identifying variations by examining similarities and differences between bacterial genomes over time. A way forward seems to be the development of a genome-wide gene-by-gene analysis tool. To this end, two approaches can be used. The first approach would involve an extended MLST (eMLST). However, instead of the traditional MLST based on seven genes, the eMLST method would be based on the whole core genome including all genes present in all isolates of a species. An allelic profile produced by eMLST would then be composed of hundreds to thousands of different alleles depending on the genome size of the investigated species. A second 'pan-genome approach' would use the full complement of genes in a species, including the core genome, the dispensable genome that represents a pool of genetic material that may be

found in a variable number of isolates within this species, and the unique genes specific to single strains of the species. In this approach, the relatedness of isolates would be measured by the presence or absence of genes across all genomes within a species. Such core- and pan-genome approaches will be endowed with a much higher discriminatory power than that of the traditional MLST, allowing the discrimination of very closely related isolates. However, to use these approaches for bacterial typing, comparative genomics must first determine the core, dispensable and unique genes among bacterial genomes at the species level. This process can be greatly facilitated by the Bacterial Isolate Genome Sequence Database (BIGSdb) comparator, and the software implemented within the web accessible PubMLST database (<http://pubmlst.org/software/database/bigfdb/>), which was created to store and compare sequence data for bacterial isolates [97]. Any number of sequences, from a single sequence read to whole genome data generated from NGS technologies, can be linked to an unlimited number of bacterial sequences. Within BIGSdb, large numbers of loci can be defined and allelic profiles for each bacterial isolate can be determined with levels of discrimination chosen on the basis of the question being asked. In this way, WGS can probably replace MLST and other typing methods currently in use. As soon as the cost of WGS comes further down and it becomes possible to perform the sequencing and analysis in <24 hours, the method will be highly useful for real-time outbreak surveillance and will likely take over as the first line surveillance typing method in any setting.

Although most typing approaches were developed to detect the presence or absence of genetic polymorphisms inside protein-encoding ORF sequences, important differences in nucleotide sequences between different bacterial strains of a species can also be observed in intergenic regions. In Europe, the predominant method for *Clostridium difficile* typing is PCR-ribotyping, which requires the PCR amplification of the intergenic space region between the 16S and 23S ribosomal RNA genes. This method yields an appropriate grouping of isolates with identical PFGE pulsotypes and has an excellent discriminatory power for isolates with different PFGE pulsotypes [98]. This supports the view that the analysis of DNA polymorphisms in intergenic regions by WGS may provide truly valuable epidemiological insights.

The genetic relatedness among bacterial isolates can also be determined by examining the genome sequence as a whole. In contrast to conventional molecular typing methods, WGS has the potential to compare different genomes with a single-nucleotide resolution. This would allow an accurate characterisation of transmission events and outbreaks. However, translating this potential into routine practice will involve extensive investigations. Methods based on SNPs permit a detailed, targeted analysis of variations within related organisms. Very recently, Köser and colleagues [91]

reported a clinically meaningful application of SNPs analysis involving the rapid high-throughput sequencing of MRSA isolates recovered from a putative outbreak in a neonatal intensive care unit. The whole genome SNPs analysis identified the isolates associated with an outbreak, and clearly separated them from other non-outbreak isolates. However, one outbreak isolate showed a higher number of SNPs than the other outbreak isolates, which highlights the difficulty in applying a simple cut-off for differences in the identified SNPs of isolates in an outbreak setting. Therefore, additional investigations and comparisons are needed to develop a strategy for automated data interpretation of an outbreak situation in clinical practice.

Interestingly, the '100K Genome Project', which is an initiative of the US Food and Drug Administration (FDA), Agilent, the University of California at Davis, and other federal and private partners, is aimed at the sequencing of 100,000 genomes of at least 100,000 food-borne pathogens over the next five years (<http://100kgenome.vetmed.ucdavis.edu>). The knowledge that is to be derived from this enormous effort will be extremely useful for epidemiological surveillance, not only due to the specific genomic information that will facilitate detailed comparisons between different bacterial isolates, but also because the data will serve as a knowledge base for the development of new pathogen detection and typing assays for outbreak investigations.

In addition to traditional epidemiological applications, WGS can also be effective for defining phenotypic characteristics, such as the virulence or antibiotic resistance of a particular pathogen [99]. First attempts to create an artificial 'resistome' of antibiotic resistance genes were already successful, as demonstrated by a comparison of genome-based predictions to the results of phenotypic susceptibility testing [91]. Similarly, based on the WGS data a potential 'toxome' was established, consisting of all toxin genes [91]. Accordingly, WGS can potentially be used to support or replace the classical determination of bacterial serotypes as it allows the detection of genes critical for the expression of particular serotype-specific antigens. However, a note of caution is in place, since the genome sequence does as yet neither allow an accurate prediction of the potentially conditional expression of particular genes, nor their expression level. This is critically underscored by proteomics studies on the cell surface and exoproteomes of different isolates of *S. aureus*, which revealed high degrees of variation in the expression of particular proteins, including known virulence factors [100-102]. Lastly, genome sequences will be also used to search for genetic markers, such as the presence or absence of a gene or an amino acid substitution in a protein, which can then be linked with an exclusive or higher occurrence in a disease, or associated with disease severity and virulence.

## Conclusions

In recent years, we have witnessed substantial technical improvements in existing approaches for the typing of bacterial isolates, and completely new technologies have emerged that will substantially impact on the way pathogenic microorganisms can be defined and distinguished in the near future. This has involved major efforts towards the automation of these typing methods, the improvement of their resolution and throughput, and the design of adequate bioinformatics tools. The steadily increasing number of genotyping databases containing DNA sequences and DNA microarray profiles now allows easier and faster inter-laboratory comparisons, retrospective analyses and long-term epidemiological surveillance of bacterial infections. Unfortunately, there is currently no single ideal typing method available, and each genotyping approach has various advantages and disadvantages. Therefore, depending on the setting (local, national or international), one or more different typing methods need to be applied. If speed is important for containing a local disease outbreak, a PCR-based method with high discriminatory power, such as MLVF and/or DiversiLab, may work well for characterisation of the isolates. However, if an outbreak of bacterial disease is disseminated among various geographical locations, a more robust typing approach, such as PFGE, will be needed to allow reliable comparison of the results obtained in different laboratories. Notably, some of the newer methods, such as MLVA, SLST, MLST, SNP or DNA microarray analysis, allow the typing of isolates equally well as the gold standard PFGE, and urgently needed results can be obtained in shorter periods of time. On the other hand, these newer methods also have certain drawbacks, including the need for highly trained staff and expensive equipment, such as automated DNA sequencers or scanners. Therefore, it is much easier to replace traditional methods with newer ones at the local level than in large national or international surveillance networks where all laboratories (with different staff and budgets) must implement the same new typing method and train all participants in its standardised application. It is important to realise that a newly introduced method must be very well validated by different independent laboratories to determine its typing potential, and this process takes years rather than months. A new method must also implement a specific unambiguous nomenclature, which needs to be developed and improved during the validation process. Accordingly, the replacement of an old well- and widely established method with a new one must be conducted gradually to avoid the loss of precious historic information generated over many years. This is underscored by the continued use of PFGE which, for example, has remained the preferred typing method in the PulseNet network for surveillance and investigation of food-borne outbreaks for over 15 years ([www.cdc.gov/pulsenet/](http://www.cdc.gov/pulsenet/)). Moreover, if a surveillance network addresses different bacterial species, it is also very convenient if the same standardised typing platform can be used for all these species. This

is another reason why PFGE is likely to remain a preferred method in PulseNet. Notably, because different typing methods are usually based on the detection of different genomic target sequences, strain variations detected with one method may remain undetected when applying another approach. Therefore, in certain situations, the combined use of several different typing methods may lead to a more precise discrimination of bacterial isolates than the use of a single method. A completely unambiguous typing of different bacterial isolates can be achieved by WGS, as this technology has the potential to resolve single base differences between two genomes. WGS thus promises to deliver high-resolution genomic epidemiology as the ultimate method for bacterial typing. However, it is presently difficult to estimate when exactly this approach will become the norm in routine laboratories. In fact, we do not anticipate that WGS can completely replace other typing systems in the near future. Compared with many conventional methods, WGS is still not a rapid and cost-effective approach. Nevertheless, recent technical improvements as well as cost reductions suggest that, in industrialised countries, WGS will gradually become a primary typing tool in routine use. Especially, bio-informatic solutions will be necessary to extract rapidly information from WGS that is important for clinical microbiology, infection control and public health. Therefore, a common web-based database will be necessary in order to have on the one side quantifiable quality control of the enormous amount of sequencing data, and to have on the other side a growing worldwide WGS-reference database. In less-resourced countries, due to limited financial resources, the well-established conventional methods like PFGE or PCR-based typing systems will probably prevail in routine laboratories in the coming decade, although these countries may then rapidly adopt WGS once it is more affordable and practical to use. In this respect, it is however important to bear in mind that all sequence-based typing methods will produce - already today - the data sets that will also be readable by the next generation, because they are based on the universal genetic code. Moreover, the challenge is to correlate continuously increasing genome sequence information with phenotypic characteristics of bacterial isolates and to make this data publically available via the Internet, thereby warranting that these achievements will be further put to clinical use not only in industrialised countries but also in less-resourced countries. Finally, the data produced by WGS will be invaluable for the development of new typing strategies and the optimisation of traditional typing methods, such as the PCR- and microarray-based approaches presented in this review.

### Acknowledgments

This work was supported by the Interreg IVA-funded projects EurSafety Health-net (III-1-02=73) and SafeGuard (III-2-03=025), part of a Dutch-German cross-border network supported by the European Commission, the German Federal

States of Nordrhein-Westfalen and Niedersachsen, and the Dutch provinces of Overijssel, Gelderland, and Limburg.

JMvD acknowledges financial support through the Top Institute Pharma projects T4-213 and T4-502.

### References

1. Struelens MJ. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin Microbiol Infect.* 1996;2(1):2-11.
2. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect.* 2007;13 Suppl 3:1-46.
3. Arbeit RD, Arthur M, Dunn R, Kim C, Selander RK, Goldstein R. Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. *J Infect Dis.* 1990;161(2):230-5.
4. Gordillo ME, Singh KV, Baker CJ, Murray BE. Typing of group B streptococci: comparison of pulsed-field gel electrophoresis and conventional electrophoresis. *J Clin Microbiol.* 1993;31(6):1430-4.
5. Prevost G, Pottecher B, Dahlet M, Bientz M, Mantz JM, Piemont Y. Pulsed field gel electrophoresis as a new epidemiological tool for monitoring methicillin-resistant *Staphylococcus aureus* in an intensive care unit. *J Hosp Infect.* 1991;17(4):255-69.
6. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol.* 1995;33(9):2233-9.
7. Tosh PK, Disbot M, Duffy JM, Boom ML, Heseltine G, Srinivasan A, et al. Outbreak of *Pseudomonas aeruginosa* surgical site infections after arthroscopic procedures: Texas, 2009. *Infect Control Hosp Epidemiol.* 2011;32(12):1179-86.
8. Yu F, Ying Q, Chen C, Li T, Ding B, Liu Y, et al. Outbreak of pulmonary infection caused by *Klebsiella pneumoniae* isolates harbouring blaIMP-4 and blaDHA-1 in a neonatal intensive care unit in China. *J Med Microbiol.* 2012;61(Pt 7):984-9.
9. McDougal LK, Steward CD, Killgore GE, Chaitram JM, McAllister SK, Tenover FC. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. *J Clin Microbiol.* 2003;41(11):5113-20.
10. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis.* 2001;7(3):382-9.
11. Murchan S, Kaufmann ME, Deplano A, de Ryck R, Struelens M, Zinn CE, et al. Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol.* 2003;41(4):1574-85.
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(7):866-75.
13. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 1995;23(21):4407-14.
14. Mortimer P, Arnold C. AFLP: last word in microbial genotyping? *J Med Microbiol.* 2001;50(5):393-5.
15. Zhao S, Mitchell SE, Meng J, Kresovich S, Doyle MP, Dean RE, et al. Genomic typing of *Escherichia coli* O157:H7 by semi-automated fluorescent AFLP analysis. *Microbes Infect.* 2000;2(2):107-13.
16. Duim B, Wassenaar TM, Rigter A, Wagenaar J. High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Appl Environ Microbiol.* 1999;65(6):2369-75.
17. Lanini S, D'Arezzo S, Puro V, Martini L, Imperi F, Piselli P, et al. Molecular epidemiology of a *Pseudomonas aeruginosa* hospital outbreak driven by a contaminated disinfectant-soap dispenser. *PLoS one.* 2011;6(2):e17064.

18. Chang HL, Tang CH, Hsu YM, Wan L, Chang YF, Lin CT, et al. Nosocomial outbreak of infection with multidrug-resistant *Acinetobacter baumannii* in a medical center in Taiwan. *Infect Control Hosp Epidemiol*. 2009;30(1):34-8.
19. Li W, Raoult D, Fournier PE. Bacterial strain typing in the genomic era. *FEMS Microbiol Rev*. 2009;33(5):892-916.
20. Versalovic J, Schneider M, de Bruijn FJ, Lupski JR. Genomic fingerprinting of bacteria using the repetitive sequence-based polymerase chain reaction. *Methods Mol Cell Biol*. 1994;5(1):25-40.
21. Sabat A, Malachowa N, Miedzobrodzki J, Hryniewicz W. Comparison of PCR-based methods for typing *Staphylococcus aureus* isolates. *J Clin Microbiol*. 2006;44(10):3804-7.
22. Wilson MK, Lane AB, Law BF, Miller WG, Joens LA, Konkel ME, et al. Analysis of the pan genome of *Campylobacter jejuni* isolates recovered from poultry by pulsed-field gel electrophoresis, multilocus sequence typing (MLST), and repetitive sequence polymerase chain reaction (rep-PCR) reveals different discriminatory capabilities. *Microb Ecol*. 2009;58(4):843-55.
23. Healy M, Huong J, Bittner T, Lising M, Frye S, Raza S, et al. Microbial DNA typing by automated repetitive-sequence-based PCR. *J Clin Microbiol*. 2005;43(1):199-207.
24. Deplano A, Denis O, Rodriguez-Villalobos H, De Ryck R, Struelens MJ, Hallin M. Controlled performance evaluation of the DiversiLab repetitive-sequence-based genotyping system for typing multidrug-resistant health care-associated bacterial pathogens. *J Clin Microbiol*. 2011;49(10):3616-20.
25. Fluit AC, Terlingen AM, Andriessen L, Ikawaty R, van Mansfeld R, Top J, et al. Evaluation of the DiversiLab system for detection of hospital outbreaks of infections by different bacterial species. *J Clin Microbiol*. 2010;48(11):3979-89.
26. Overdeest IT, Willemsen I, Elberts S, Verhulst C, Rijnsburger M, Savelkoul P, et al. Evaluation of the DiversiLab typing method in a multicenter study assessing horizontal spread of highly resistant gram-negative rods. *J Clin Microbiol*. 2011;49(10):3551-4.
27. Babouee B, Frei R, Schultheiss E, Widmer AF, Goldenberger D. Comparison of the DiversiLab repetitive element PCR system with spa typing and pulsed-field gel electrophoresis for clonal characterization of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol*. 2011;49(4):1549-55.
28. Sabat A, Krzyszton-Russjan J, Strzalka W, Filipek R, Kosowska K, Hryniewicz W, et al. New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. *J Clin Microbiol*. 2003;41(4):1801-4.
29. Cavanagh JP, Klingenberg C, Hanssen AM, Fredheim EA, Francois P, Schrenzel J, et al. Core genome conservation of *Staphylococcus haemolyticus* limits sequence based population structure analysis. *J Microbiol Methods*. 2012;89(3):159-66.
30. Francois P, Huyghe A, Charbonnier Y, Bento M, Herzig S, Topolski I, et al. Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of *Staphylococcus aureus* isolates. *J Clin Microbiol*. 2005;43(7):3346-55.
31. Fillo S, Giordani F, Anniballi F, Gorge O, Ramisse V, Vergnaud G, et al. Clostridium botulinum group I strain genotyping by 15-locus multilocus variable-number tandem-repeat analysis. *J Clin Microbiol*. 2011;49(12):4252-63.
32. Sabat AJ, Chlebowicz MA, Grundmann H, Arends JP, Kampinga G, Meessen NE, et al. Microfluidic-chip-based multiple-locus variable-number tandem-repeat fingerprinting with new primer sets for methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol*. 2012;50(7):2255-62.
33. Elberse KE, Nunes S, Sa-Leao R, van der Heide HG, Schouls LM. Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. *PLoS one*. 2011;6(5):e19668.
34. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, van Santen-Verheul MG, et al. Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. *PLoS one*. 2009;4(4):e5082.
35. Amonsin A, Li LL, Zhang Q, Bannantine JP, Motiwala AS, Sreevatsan S, et al. Multilocus short sequence repeat sequencing approach for differentiating among *Mycobacterium avium* subsp. *paratuberculosis* strains. *J Clin Microbiol*. 2004;42(4):1694-702.
36. Danin-Poleg Y, Cohen LA, Gancz H, Broza YY, Goldshmidt H, Malul E, et al. *Vibrio cholerae* strain typing and phylogeny study based on simple sequence repeats. *J Clin Microbiol*. 2007;45(3):736-46.
37. Visca P, D'Arezzo S, Ramisse F, Gelfand Y, Benson G, Vergnaud G, et al. Investigation of the population structure of *Legionella pneumophila* by analysis of tandem repeat copy number and internal sequence variation. *Microbiology*. 2011;157(Pt 9):2582-94.
38. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infect Dis*. 2009;9(10):611-6.
39. Beall B, Facklam R, Thompson T. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol*. 1996;34(4):953-8.
40. Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, et al. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol*. 2006;44(7):2524-32.
41. Bessen DE, McGregor KF, Whatmore AM. Relationships between emm and multilocus sequence types within a global collection of *Streptococcus pyogenes*. *BMC Microbiol*. 2008;8:59.
42. Mellmann A, Mosters J, Bartelt E, Roggentin P, Ammon A, Friedrich AW, et al. Sequence-based typing of *flaB* is a more stable screening tool than typing of *flaA* for monitoring of *Campylobacter* populations. *J Clin Microbiol*. 2004;42(10):4840-2.
43. Niederer L, Kuhnert P, Egger R, Buttner S, Hachler H, Korczak BM. Genotypes and antibiotic resistances of *Campylobacter jejuni* and *Campylobacter coli* isolates from domestic and travel-associated human cases. *Appl Environ Microbiol*. 2012;78(1):288-91.
44. Frenay HM, Bunschoten AE, Schouls LM, van Leeuwen WJ, Vandenbroucke-Grauls CM, Verhoef J, et al. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis*. 1996;15(1):60-4.
45. Luczak-Kadlubowska A, Sabat A, Tambic-Andrasevic A, Payerl-Pal M, Krzyszton-Russjan J, Hryniewicz W. Usefulness of multiple-locus VNTR fingerprinting in detection of clonality of community- and hospital-acquired *Staphylococcus aureus* isolates. *Antonie Van Leeuwenhoek*. 2008;94(4):543-53.
46. Malachowa N, Sabat A, Gniadkowski M, Krzyszton-Russjan J, Empel J, Miedzobrodzki J, et al. Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, spa typing, and multilocus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. *J Clin Microbiol*. 2005;43(7):3095-100.
47. Deurenberg RH, Nulens E, Valvatne H, Sebastian S, Driessen C, Craeghs J, et al. Cross-border dissemination of methicillin-resistant *Staphylococcus aureus*, Euregio Meuse-Rhin region. *Emerg Infect Dis*. 2009;15(5):727-34.
48. Friedrich AW, Daniels-Haardt I, Kock R, Verhoeven F, Mellmann A, Harmsen D, et al. EUREGIO MRSA-net Twente/Munsterland--a Dutch-German cross-border network for the prevention and control of infections caused by methicillin-resistant *Staphylococcus aureus*. *Euro Surveill*. 2008;13(35):pii=18965. Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18965>
49. Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, Friedrich AW. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. *PLoS Med*. 2010;7(1):e1000215.
50. Hallin M, Deplano A, Denis O, De Mendonca R, De Ryck R, Struelens MJ. Validation of pulsed-field gel electrophoresis and spa typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J Clin Microbiol*. 2007;45(1):127-33.
51. Harmsen D, Claus H, Witte W, Rothganger J, Claus H, Turnwald D, et al. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol*. 2003;41(12):5442-8.
52. Kock R, Brakensiek L, Mellmann A, Kipp F, Henderikx M, Harmsen D, et al. Cross-border comparison of the admission prevalence and clonal structure of methicillin-resistant *Staphylococcus aureus*. *J Hosp Infect*. 2009;71(4):320-6.
53. Friedrich AW, Witte W, Harmsen D, de Lencastre H, Hryniewicz W, Scheres J, et al. SeqNet.org: a European

- laboratory network for sequence-based typing of microbial pathogens. *Euro Surveill.* 2006;11(2):pii=2874. Available from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2874>
54. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol.* 1986;51(5):873-84.
  55. 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 U S A.* 1998;95(6):3140-5.
  56. Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, et al. Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol.* 2001;39(1):14-23.
  57. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol.* 2000;38(3):1008-15.
  58. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology.* 1998;144 ( Pt 11):3049-60.
  59. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect Immun.* 2001;69(4):2416-27.
  60. Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, et al. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol.* 2003;41(5):2068-79.
  61. Homan WL, Tribe D, Poznanski S, Li M, Hogg G, Spalburg E, et al. Multilocus sequence typing scheme for *Enterococcus faecium*. *J Clin Microbiol.* 2002;40(6):1963-71.
  62. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, Kunst F, et al. Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol.* 2003;41(6):2530-6.
  63. King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, et al. Development of a multilocus sequence typing scheme for the pig pathogen *Streptococcus suis*: identification of virulent clones and potential capsular serotype exchange. *J Clin Microbiol.* 2002;40(10):3671-80.
  64. 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(4):1623-36.
  65. Salcedo C, Arreaza L, Alcalá B, de la Fuente L, Vazquez JA. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J Clin Microbiol.* 2003;41(2):757-62.
  66. Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 2003;11(10):479-87.
  67. Martin IM, Ison CA, Aanensen DM, Fenton KA, Spratt BG. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. *J Infect Dis.* 2004;189(8):1497-505.
  68. Tankouo-Sandjong B, Sessitsch A, Liebana E, Kornschöber C, Allerberger F, Hachler H, et al. MLST-v, multilocus sequence typing based on virulence genes, for molecular typing of *Salmonella enterica* subsp. *enterica* serovars. *J Microbiol Methods.* 2007;69(1):23-36.
  69. Zhang W, Jayarao BM, Knabel SJ. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Appl Environ Microbiol.* 2004;70(2):913-20.
  70. Teh CS, Chua KH, Thong KL. Genetic variation analysis of *Vibrio cholerae* using multilocus sequencing typing and multi-virulence locus sequencing typing. *Infect Genet Evol.* 2011;11(5):1121-8.
  71. Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smith P, Ribot EM, et al. Subtyping *Salmonella enterica* serovar enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). *Appl Environ Microbiol.* 2011;77(13):4520-6.
  72. Verghese B, Schwalm ND, 3rd, Dudley EG, Knabel SJ. A combined multi-virulence-locus sequence typing and *Staphylococcal Cassette Chromosome mec* typing scheme possesses enhanced discriminatory power for genotyping MRSA. *Infect Genet Evol.* 2012;12(8):1816-21.
  73. McCarthy AJ, Breathnach AS, Lindsay JA. Detection of mobile-genetic-element variation between colonizing and infecting hospital-associated methicillin-resistant *Staphylococcus aureus* isolates. *J Clin Microbiol.* 2012;50(3):1073-5.
  74. McCarthy AJ, Lindsay JA. The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol.* 2012;12:104.
  75. Lindsay JA, Moore CE, Day NP, Peacock SJ, Witney AA, Stabler RA, et al. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *J Bacteriol.* 2006;188(2):669-76.
  76. Jackson SA, Kotewicz ML, Patel IR, Lacher DW, Gangiredla J, Elkins CA. Rapid genomic-scale analysis of *Escherichia coli* O104:H4 by using high-resolution alternative methods to next-generation sequencing. *Appl Environ Microbiol.* 2012;78(5):1601-5.
  77. Monecke S, Coombs G, Shore AC, Coleman DC, Akpaka P, Borg M, et al. A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. *PLoS one.* 2011;6(4):e17936.
  78. Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehrlich R, Hachler H. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. *J Clin Microbiol.* 2007;45(2):370-9.
  79. Braun SD, Ziegler A, Methner U, Slickers P, Keiling S, Monecke S, et al. Fast DNA Serotyping and Antimicrobial Resistance Gene Determination of *Salmonella enterica* with an Oligonucleotide Microarray-Based Assay. *PLoS one.* 2012;7(10):e46489.
  80. Lim A, Dimalanta ET, Potamou K, Yen G, Apodoca J, Tao C, et al. Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. *Genome Res.* 2001;11(9):1584-93.
  81. Aston C, Mishra B, Schwartz DC. Optical mapping and its potential for large-scale sequencing projects. *Trends Biotechnol.* 1999;17(7):297-302.
  82. Johnson PD, Ballard SA, Grabsch EA, Stinear TP, Seemann T, Young HL, et al. A sustained hospital outbreak of vancomycin-resistant *Enterococcus faecium* bacteremia due to emergence of vanB E. faecium sequence type 203. *J Infect Dis.* 2010;202(8):1278-86.
  83. Kotewicz ML, Mammel MK, LeClerc JE, Cebula TA. Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology.* 2008;154(Pt 11):3518-28.
  84. Petersen RF, Litrup E, Larsson JT, Torpdahl M, Sorensen G, Muller L, et al. Molecular characterization of *Salmonella* Typhimurium highly successful outbreak strains. *Foodborne Pathog Dis.* 2011;8(6):655-61.
  85. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS one.* 2011;6(7):e22751.
  86. Ben Zakour NL, Venturini C, Beatson SA, Walker MJ. Analysis of a *Streptococcus pyogenes* puerperal sepsis cluster by use of whole-genome sequencing. *J Clin Microbiol.* 2012;50(7):2224-8.
  87. 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(1):33-42.
  88. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, et al. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci U S A.* 2012;109(8):3065-70.
  89. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS one.* 2012;7(11):e47768.
  90. Vernet G, Saha S, Satzke C, Burgess DH, Alderson M, Maisonneuve JF, et al. Laboratory-based diagnosis of pneumococcal pneumonia: state of the art and unmet needs. *Clin Microbiol Infect.* 2011;17 Suppl 3:1-13.
  91. Koser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med.* 2012;366(24):2267-75.
  92. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol.* 2012;50(4):1355-61.
  93. Askar M, Faber MS, Frank C, Bernard H, Gilsdorf A, Fruth A, et al. Update on the ongoing outbreak of haemolytic uraemic syndrome due to Shiga toxin-producing *Escherichia coli* (STEC) serotype O104, Germany, May 2011. *Euro Surveill.* 2011;16(22):pii=19883. Available

from: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19883>

94. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, et al. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med*. 2011;365(19):1771-80.
95. Mellmann A, Bielaszewska M, Kock R, Friedrich AW, Fruth A, Middendorf B, et al. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis*. 2008;14(8):1287-90.
96. Hao W, Allen VG, Jamieson FB, Low DE, Alexander DC. Phylogenetic incongruence in *E. coli* O104: understanding the evolutionary relationships of emerging pathogens in the face of homologous recombination. *PLoS one*. 2012;7(4):e33971.
97. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010;11:595.
98. Westblade LF, Chamberland RR, Maccannell D, Collins R, Dubberke ER, Dunne WM, Jr., et al. Development and Evaluation of a Novel, Semi-Automated *Clostridium difficile* Typing Platform. *J Clin Microbiol*. 2012.
99. Billal DS, Feng J, Leprohon P, Legare D, Ouellette M. Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations. *BMC Genomics*. 2011;12:512.
100. Sibbald MJ, Ziebandt AK, Engelmann S, Hecker M, de Jong A, Harmsen HJ, et al. Mapping the pathways to staphylococcal pathogenesis by comparative secretomics. *Microbiol Mol Biol Rev*. 2006;70(3):755-88.
101. Ziebandt AK, Kusch H, Degner M, Jaglitz S, Sibbald MJ, Arends JP, et al. Proteomics uncovers extreme heterogeneity in the *Staphylococcus aureus* exoproteome due to genomic plasticity and variant gene regulation. *Proteomics*. 2010;10(8):1634-44.
102. Dreisbach A, Hempel K, Buist G, Hecker M, Becher D, van Dijl JM. Profiling the surfacome of *Staphylococcus aureus*. *Proteomics*. 2010;10(17):3082-96.

