



Application of next generation sequencing in clinical microbiology and infection prevention



Ruud H. Deurenberg^{a,1}, Erik Bathoorn^{a,1}, Monika A. Chlebowicz^{a,1}, Natacha Couto^{a,1}, Mithila Ferdous^{a,1}, Silvia García-Cobos^{a,1}, Anna M.D. Kooistra-Smid^{a,b,1}, Erwin C. Raangs^{a,1}, Sigrid Rosema^{a,1}, Alida C.M. Veloo^{a,1}, Kai Zhou^{c,1}, Alexander W. Friedrich^a, John W.A. Rossen^{a,*}

^a Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, The Netherlands

^b Certe, Department of Medical Microbiology, Groningen, The Netherlands

^c State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital of Medicine School, Zhejiang University, Hangzhou, China

ARTICLE INFO

Article history:

Received 31 October 2016

Received in revised form

27 December 2016

Accepted 28 December 2016

Available online 29 December 2016

Keywords:

Infection prevention

Ion PGM™

Clinical microbiology

MiSeq

Next generation sequencing

Whole genome sequencing

ABSTRACT

Current molecular diagnostics of human pathogens provide limited information that is often not sufficient for outbreak and transmission investigation. Next generation sequencing (NGS) determines the DNA sequence of a complete bacterial genome in a single sequence run, and from these data, information on resistance and virulence, as well as information for typing is obtained, useful for outbreak investigation. The obtained genome data can be further used for the development of an outbreak-specific screening test. In this review, a general introduction to NGS is presented, including the library preparation and the major characteristics of the most common NGS platforms, such as the MiSeq (Illumina) and the Ion PGM™ (ThermoFisher). An overview of the software used for NGS data analyses used at the medical microbiology diagnostic laboratory in the University Medical Center Groningen in The Netherlands is given. Furthermore, applications of NGS in the clinical setting are described, such as outbreak management, molecular case finding, characterization and surveillance of pathogens, rapid identification of bacteria using the 16S-23S rRNA region, taxonomy, metagenomics approaches on clinical samples, and the determination of the transmission of zoonotic micro-organisms from animals to humans. Finally, we share our vision on the use of NGS in personalised microbiology in the near future, pointing out specific requirements.

© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Identification and characterization of micro-organisms that cause infections are crucial for successful treatment, recovery and safety of patients. However, not every bacterial species can be successfully cultured in the diagnostic laboratory, and the available molecular tests are unable to detect emerging genetic features in successfully evolving pathogens that spread in humans, animals and the environment. Unrecognized pathogens can easily cause hospital outbreaks, putting patients at risk during their hospital admissions.

During the last two decades, molecular diagnostic methods have experienced a rapid development and played an increasingly

important role in medical microbiology laboratories (Buchan and Ledeboer, 2014). These methods have reduced the turnaround time from receiving the sample to the final result, and made it possible to detect non-cultivable pathogens. However, molecular methods need a priori knowledge of the likely pathogenic species that could be present in the sample. One of the molecular methods used in medical microbiology laboratories is the sequence analyses of genes or the whole genome of pathogens.

Sequence analyses can be used to answer different diagnostic questions, such as the genetic relationship of either bacteria or viruses, the detection of mutations in viral or bacterial genomes leading to resistance against antivirals or antibiotics, identification of fungi through sequence analyses of the 18S ribosomal deoxyribonucleic acid (rDNA) of the internal transcribed spacer (ITS) region and identification of bacteria through sequence analyses of the 16S rDNA (Bush, 2013; Deurenberg and Stobberingh, 2008; Liu et al., 2012; Reiss et al., 2000). In general, Sanger sequencing is

* Corresponding author.

E-mail address: j.w.a.rossen@rug.nl (J.W.A. Rossen).

¹ These authors have contributed equally to this work.

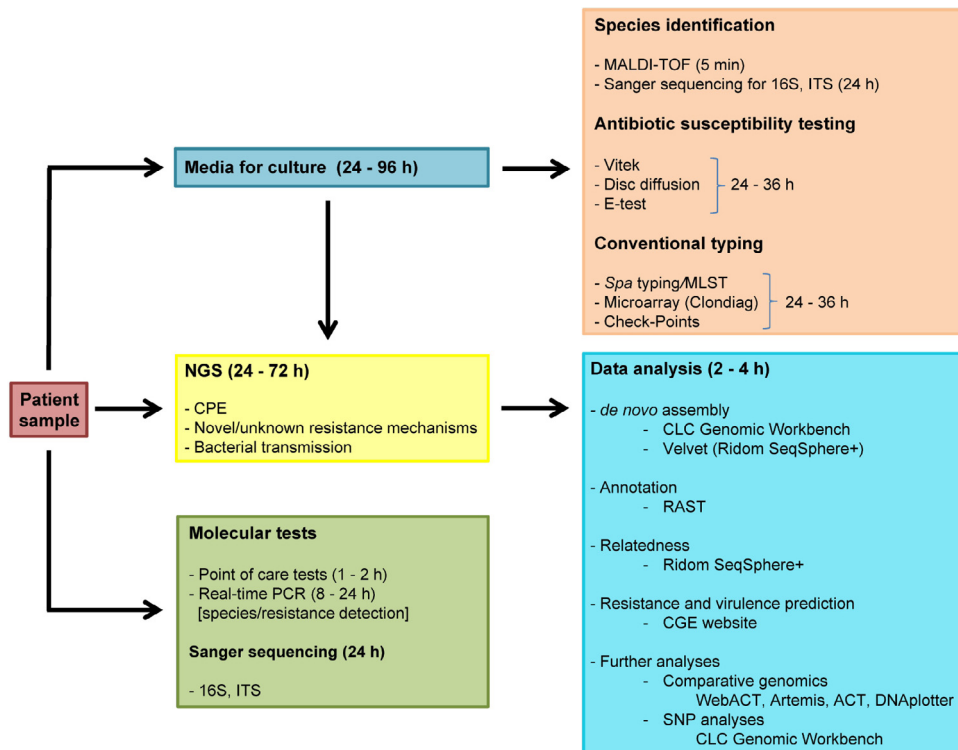


Fig. 1. A schematic overview of the general workflow of diagnostic procedures including NGS in our laboratory.

used for this, preceded by amplification of each gene or genomic region using specific primers. The same method can be applied for the identification of pathogens in clinical material. However, this approach becomes problematic when clinical material is more complex and contains multiple species, such as faecal samples. In such cases, results obtained by Sanger sequencing are not reliable and make it hard or even impossible to identify specific pathogens. Furthermore, the cost of Sanger sequencing for these tasks is high, and the turnaround time is long.

The University Medical Center Groningen (UMCG) is one of the largest university hospitals in The Netherlands with 1339 beds and more than 12,000 employees. The clinical microbiology diagnostic laboratory at the UMCG receives around 5750 samples per year for detailed molecular analysis, of which approximately 1500 are subjected to next generation sequencing (NGS) using two Illumina MiSeq and one Life Technologies Ion PGM™ sequencers. NGS was introduced for routine diagnostics in 2014, and the majority of indications are outbreak investigation and genotyping of highly resistant micro-organisms. NGS is requested by clinical microbiologists or infectious disease specialists in collaboration with molecular microbiologists and infection control practitioners.

2. Next generation sequencing

NGS allows sequencing of the whole genome of numerous pathogens in one sequence run, either from bacterial isolates of (different) patients, or from multiple species present in patient material from one individual (metagenomics). Both the investment- and the running costs of NGS have decreased dramatically during the last decade (Dark, 2013; Sboner et al., 2011). A great advantage of NGS is that, in contrast to Sanger sequencing, a single protocol can be used for all pathogens for both identification and typing applications. Therefore, this technology has been proven to be useful in medical microbiology laboratories and for infection prevention measures (Zhou et al., 2016). A schematic overview of

the general workflow used for NGS analyses at the UMCG is shown in Fig. 1.

For NGS, there is no need for target specific primers, which are needed for Sanger sequencing. In a single run, the whole genome of a pathogen is sequenced at random. Before sequencing, fragmentation of the genome is performed, since the maximum length a benchtop sequencer can sequence varies between 100 and 1000 bases and thus the genome cannot be sequenced in one part (Junemann et al., 2013; Loman et al., 2012). An exception to this are the third generation of sequencers, such as the MinION (Oxford Nanopore) and the Sequel (Pacific Biosciences), which can generate larger fragments (more than 200 kb). However, these sequencers are not yet used in the clinical microbiology laboratory, due to their lack of affordability, the lower quality of the sequences, and the low throughput. Therefore, NGS still requires the preparation of libraries, in which fragments of DNA or RNA are fused to adapters and barcodes to distinguish the DNA of the sequenced isolates after sequencing, followed by a clonal amplification, normalization and sequencing. For this, a robust preparation of the libraries, which contains a representative source of the DNA or RNA of the genome under investigation, is needed (Head et al., 2014).

Fragmentation can be performed in several ways, either mechanical, using, e.g., the Adaptive Focused Acoustics (AFA) technology from Covaris, followed by adaptor ligation, or enzymatic, such as with transposons as used in the Nextera XT Library Preparation kit from Illumina. This method has the advantage that fragmentation and fusion of the adaptors to the DNA or RNA fragments are performed in one step, which makes it easier to automate it. Besides that, less input DNA is needed. Mechanical fragmentation has the advantage that the generation of the appropriate fragment length is less influenced by factors present in the sample that inhibit the enzymes used during the library preparation, and is therefore very suitable for library preparations of direct sample material, such as biopsies and faeces samples (Head et al., 2014).

NGS libraries may contain errors that decrease the data quality, and thus can disrupt the data interpretation. Detailed knowledge of

Table 1
Properties of current NGS platforms.

| Company | Equipment | Output/run (Gb) | Maximum read length (bp) | Reads (x10 ⁶) | Running time |
|----------------------------------|------------|-----------------|--------------------------|---------------------------|--------------|
| Illumina | MiniSeq | 0.6–7.5 | 2 × 150 | 25 | 4–24 h |
| Illumina | Miseq | 0.3–15 | 2 × 300 | 25 | 5–55 h |
| Illumina | NextSeq | 20–120 | 2 × 150 | 130/400 | 12–30 h |
| Illumina | HiSeq 3000 | 125–700 | 2 × 150 | 2500 | <1–3.5 days |
| ThermoFisher | Ion PGM™ | 0.03–2 | 200–400 | 0.4–5.5 | 2–7 h |
| ThermoFisher | Ion 5S™ | 0.6–15 | 200–400 | 3–80 | 2.5–4 h |
| ThermoFisher | Ion 5S™ XL | 0.6–15 | 200–400 | 3–80 | <24 h |
| Oxford Nanopore | MinION | 21–42 | 230,000–300,000 | 2.2–4.4 | 1 min–48 h |
| Pacific Biosciences ^a | Sequel | 0.75–1.25 | >20,000 | 370,000 | 30 min–6 h |
| Pacific Biosciences ^a | RSII | 0.5–1 | >20,000 | 55,000 | 30 min–4 h |

^a The Pacific Biosciences data are per smart cell; both the Sequel and the RSII can run 1–16 smart cells in one run.

the kind of errors is important to find ways to avoid the introduction of such errors and for the correct interpretation of the NGS data. Almost all separate steps in the sequence procedure can introduce errors. This is especially true with RNA sequencing that is technically more challenging compared to DNA sequencing (Junemann et al., 2013; Loman et al., 2012).

At the moment, a number of NGS platforms are available. The most important properties of several NGS platforms, such as output and fragment length, are presented in Table 1 (Bertelli and Greub, 2013; Dark, 2013; Junemann et al., 2013; Loman et al., 2012). The different NGS platforms use different sequencing technologies. Illumina sequencers use sequencing by synthesis of fluorescent, reversible terminators, and ThermoFisher sequencers use semiconductor sequencing that measure a change in pH during the incorporation of nucleotides. Pacific Biosciences use fluorescent nucleotides in their single molecule real-time (SMRT) technology, and Oxford Nanopore platforms use ionic current sensing, in which DNA is guided through nano-pores, thereby changing the current in a way that is specific for the type of nucleotide. Extensive information on the different NGS platforms and their method of sequencing are available on the companies' websites. Due to the technological developments, the cost of NGS decreased between 2001 and 2015, while the speed of sequencing has increased (Dark, 2013; Stoner et al., 2011).

3. Software for data analyses

The biggest challenge concerning the introduction of NGS in the clinical microbiology laboratory is the data analyses. Nonetheless, even with little knowledge of bioinformatics, it is possible to perform NGS data analyses for diagnostic purposes, using the numerous user-friendly software packages available (Edwards and Holt, 2013). However, for more in-depth analysis, scientific knowledge is required on the genomic features and the biological background of the micro-organism under investigation.

After sequencing, the sequenced fragments (reads) can be *de novo* assembled (genome assembly). Hereby, the reads are aligned against each other without the use of a reference organism. In general, the larger the fragments, the easier and more accurate the genome assembly will be. Software packages (Table 2), such as CLC Genomic Workbench (Qiagen), SPAdes and Velvet, are used in our laboratory to assemble the genomes. The genetic relationship between isolates can be investigated by using a gene-by-gene comparison using a multi-locus sequencing typing (MLST) approach, either by studying the conserved core genome (cgMLST), or the whole genome (wgMLST), which includes a set of variable accessory genes. Several software packages, such as SeqSphere (Ridom) and BioNumerics (Applied Maths, Biomérieux), or online tools, such as EnteroBase and BIGSdb (Bacterial Isolate Genome Sequence Database) (Jolley and Maiden, 2010), can be used for this approach.

Furthermore, the use of an established cgMLST scheme allows the introduction of a common nomenclature for genetically related strains (de Been et al., 2015; Kohl et al., 2014; Ruppitsch et al., 2015). At the moment, it is not clear how many alleles two genomes may differ to call them (close to being) identical. The same problem applies when comparing two genomes using single-nucleotide polymorphisms (SNP) typing (Maiden et al., 2013). However, the term genetic distance (the proportion of different alleles, calculated by dividing the number of allele differences by the total number of genes shared by two sequences) has been recently introduced, and enables unbiased comparisons for different cgMLST or wgMLST schemes as well as the definition of thresholds by studying collections of epidemiologically and non-epidemiologically related strains (Kluytmans-van den Bergh et al., 2016b). An advantage of cgMLST and wgMLST is that there is compatibility between cgMLST, wgMLST and older typing methods, since both BioNumerics and SeqSphere give the sequence type (ST) as determined by conventional MLST and the *spa* type (in case of *Staphylococcus aureus*).

Several possibilities to analyse NGS data can be found on the website of the Center for Genomic Epidemiology. For the detection of virulence- and resistance genes, VirulenceFinder and ResFinder, can be used. Alternatively, the Comprehensive Antibiotic Resistance Database (CARD) and the Virulence Factor Database (VFDB) can be used to obtain data on resistance and virulence genes. With these online tools, both the non-assembled sequence data and the assembled genome can be uploaded. However, the results obtained through these websites needs confirmation using other methods. In case of *S. aureus*, it has been reported that there is a good correlation between the presence of resistance genes and its phenotypic resistance pattern (Aanensen et al., 2016).

Further comparative genome studies are possible using Artemis Comparison Tool (ACT) (Carver et al., 2005), Artemis (Carver et al., 2012), and DNA plotter (Carver et al., 2009) from the Sanger Institute.

Specific research questions require knowledge of Unix-systems and to handle the large diversity of bioinformatics software packages available for it. Furthermore, software for the analyses of metagenomics data is available, such as the MEGAN Alignment Tool (Huson et al., 2007). In contrast to the decreased sequencing costs, the costs for data storage and data analyses have increased due the generation of large amounts of data, and the complexity of it.

4. NGS in clinical microbiology

NGS is already applied in several medical microbiology laboratories, including our laboratory at the UMCG, where it is used for outbreak management, molecular case finding, characterization and surveillance of pathogens, rapid identification of bacteria using the 16S–23S rRNA region, taxonomy, metagenomics approaches

Table 2

Software packages frequently used for NGS data analyses in our laboratory.

| Application | Software | Link | Note |
|--------------------------------------|---|---|---|
| Annotation | Prokka RAST | www.vicbioinformatics.com http://rast.nmpdr.org | |
| Assembly | BioNumerics CLC Genomic Workbench SeqSphere SPAdes Velvet | www.applied-maths.com www.clcbio.com www.ridom.de http://bioinf.spbau.ru/spades www.ebi.ac.uk/~zerbino/velvet | Commercial software Commercial software Commercial software Unix-based Unix-based |
| Data quality check | BaseSpace BioNumerics CLC Genomic Workbench FastQC | https://basespace.illumina.com www.applied-maths.com www.clcbio.com www.bioinformatics.babraham.ac.uk | Commercial software Commercial software Commercial software |
| Identification | K-merFinder NCBI BLAST | www.genomicepidemiology.org www.ncbi.nlm.nih.gov/blast | |
| Metagenomics Phylogeny | MEGAN FastTree RAxML SeqSphere SNPTree | http://ab.inf.uni-tuebingen.de/software/malt www.microbesonline.org/fasttree http://sco.h-its.org/exelixis/software.html www.ridom.de www.genomicepidemiology.org | Commercial software |
| Resistance | ARDB CARD ResFinder | https://ardb.cbc.umd.edu https://card.mcmaster.ca www.genomicepidemiology.org | |
| SNP calling | BioNumerics CLC Genomic Workbench Samtools SeqSphere | www.applied-maths.com www.clcbio.com www.htslib.org www.ridom.de | Commercial software Commercial software Commercial software |
| Typing (wgMLST) | BIGSdb BioNumerics CLC Genomic Workbench Enterobase SeqSphere | http://bigsdb.readthedocs.io www.applied-maths.com www.clcbio.com https://enterobase.warwick.ac.uk www.ridom.de | Commercial software Commercial software Commercial software |
| Virulence | VFDB VirulenceFinder | www.mgc.ac.cn/VFs www.genomicepidemiology.org | |
| Visualisation & comparative study | ACT Artemis BRIG ClustalW DNA plotter WebACT | www.sanger.ac.uk/science/tools www.sanger.ac.uk/science/tools https://sourceforge.net/projects/brig/ www.genome.jp/tools/clustalw www.sanger.ac.uk/science/tools www.webact.org | |

on clinical samples, and the determination of the transmission of zoonotic micro-organisms from animals to humans.

4.1. Outbreak management

The advantages of using whole genome sequencing (WGS)-based typing is promoting the implementation of NGS for epidemiological studies and public health investigations. It is especially important and helpful in outbreak detection and monitoring the evolution and dynamics of multi-drug resistant pathogens (ECDC, 2016). Several studies illustrated the usefulness of WGS-based typing for disclosing and tracing the dissemination of emerging pathogens. Indeed, it was used in our hospital to characterize a newly emerging CTX-M-15 producing *K. pneumoniae* clone with sequence type (ST) 1427 (Zhou et al., 2015b). In addition, the transmission of a CTX-M-15-producing ST15 *Klebsiella pneumoniae* between patients treated in a single centre and the subsequent inter-institutional spread by patient referral has been traced by genomic phylogenetic analysis (Fig. 2). The investigation allowed the early detection of a *K. pneumoniae* high-risk-clone (HiRiC) with prolonged circulation in the regional patient population (Zhou et al., 2016). Furthermore, this study showed the usefulness of a unique marker approach, in which a clone-specific PCR was developed to investigate the dissemination of the HiRiC between healthcare centres.

In addition to outbreak tracing and characterization, the use of WGS also allows the implementation of control measures to avoid the spread of resistant bacterial clones. An outbreak of a colistin-resistant carbapenemase-producing *K. pneumoniae* (KPC) with inter-institutional spread in The Netherlands, was controlled by transferring all positive residents to a separate location outside the institution, where a dedicated team cared for the patients (Weterings et al., 2015).

Apart from multidrug-resistant bacteria, WGS is also useful and applicable to characterize highly-virulent bacteria, such as shiga toxin-producing *Escherichia coli* (STEC) O104:H4. This bacterium has emerged as an important pathogen and has been responsible for large outbreaks. However, there has been little information about the evolutionary history and genomic diversity of it. Phylogenetic analysis of outbreak and non-outbreak related isolates using WGS provided an evolutionary context and revealed lineage-specific markers, indicative for selective pressure and niche adaptation (Zhou et al., 2015a). In addition, core-genome phylogenetic analysis of shiga toxin-producing Enterohaemorrhagic *E. coli* (EAEC Stx2a+) O104:H4 showed different clustering and different resistance and virulence patterns depending on the time of isolation (Ferdous et al., 2015). These studies reflect the importance of NGS as a high discriminatory power tool to differentiate between clones with specific properties and to use the obtained knowledge for patient management, infection prevention and evolutionary studies.

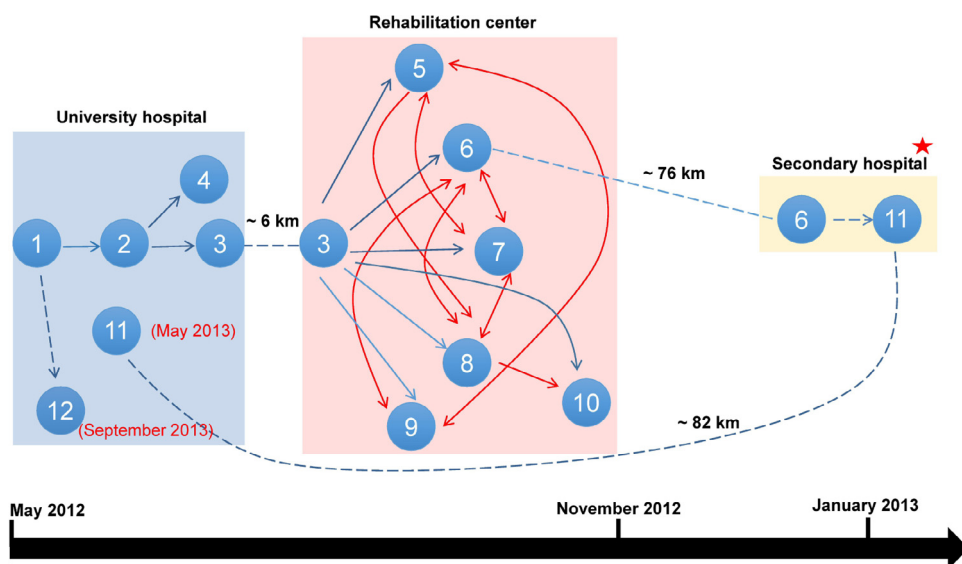


Fig. 2. The transmission route was reconstructed by epidemiological and genomic data. Each node represents a patient, and an arrow indicates a possible transmission event from one patient to another. The blue arrow with solid line represents a direct transmission event supported by both epidemiological data and genetic data, the blue arrow with dash line represents an indirect transmission (e.g. via environment) supported by epidemiological data, and the red arrow indicates the equally parsimonious transmission link which cannot be resolved by neither epidemiological data nor genetic data. The inter-institutional transfer of the patient is shown by dash lines, on which the distance between institutions is indicated. The red star represents an outbreak at a secondary hospital, but the isolates were unavailable for further research (Zhou et al., 2016). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Molecular case finding

Molecular case definitions of outbreak isolates are commonly used in outbreak investigations. NGS databases can retrospectively be searched for cases in complex and comprehensive outbreaks. This may result in detection of cases that would not have been found by traditional epidemiological investigation. In a study, a New Delhi Metallo- β -lactamase-5 (NDM-5)-producing *K. pneumoniae* ST16 strain was isolated from a Dutch patient from a long-term care facility without a recent history of travel abroad. Molecular case finding showed that the Dutch strain was clonally related to strains isolated from four patients in Denmark in 2014 (Bathoorn et al., 2015), but there were no obvious epidemiological links between the cases in the Danish and Dutch hospitals. European national surveillance centres were contacted for molecular case finding in their NGS databases, but no additional cases were detected.

Other examples of molecular case finding in NGS databases have been reported after the discovery of the plasmid resistance gene *mcr-1*, responsible for colistin resistance, in life-stock and hospitalised patients. Its introduction in European countries was investigated using a retrospectively search for the *mcr-1* gene in NGS databases. This resulted in the detection of cases in Denmark, Germany, and The Netherlands (Falgenhauer et al., 2016; Hasman et al., 2015; Kluytmans-van den Bergh et al., 2016a). In The Netherlands, more than 2000 Dutch *Enterobacteriaceae* isolates were screened within a few hours to reveal the presence of the *mcr-1* gene (Kluytmans-van den Bergh et al., 2016a). So, NGS data already present can be used to screen for the presence of new (antibiotic resistance) genes (*in silico* screening).

4.3. Characterization and surveillance of pathogens

The current routine procedure for pathogen characterization is based on a large variety of bacteriological, biochemical and molecular methods, making this procedure laborious, time-consuming, and expensive. NGS may serve as a perfect one-step tool to study a broad range of pathogen characteristics and is applicable on a wide range of pathogens (Aanensen et al., 2016; Fournier et al.,

2014; Hasman et al., 2014). Knowledge of the virulence profile of a pathogen is crucial to predict the disease severity, outcome of the infection and to allow risk assessment during the early onset of the disease. WGS has the potential to make a substantial contribution to determine the presence of virulence factors using several online tools, since it is not restricted to a specific gene (Franz et al., 2014; Laabei et al., 2014).

In a large cohort study, WGS was used for molecular characterization of STEC, resulting in a clear understanding about the population structure and genomic plasticity of STEC in the regions around the cities Groningen and Rotterdam in The Netherlands (Ferdous et al., 2016). All relevant information could be extracted *in silico* from the sequence data, including genotype, serotype, MLST profile, virulence and antibiotic resistance profiles, and the phylogenetic background to obtain the overall molecular features with a high discrimination among closely related strains. NGS allowed to characterize and compare many strains in detail within a relatively short time span. Thus, for a rapid and improved molecular epidemiological surveillance of pathogens at regional and national scale, the role of WGS is undeniable.

NGS is also helpful in detection of novel resistance genes in bacteria, both in current as well as in historical strain collections. Novel variants of antibiotic resistance genes (ARG) can be identified using NGS, and further experiments can be performed to determine if these genes are indeed responsible for the observed antibiotic resistance pattern (Nijhuis et al., 2015).

4.4. Targeted NGS of the 16S-23S rRNA cluster region for rapid bacterial identification in clinical specimen

NGS allows culture-free detection of a theoretically unlimited number of pathogens and thus provides insight in the full microbiome. Metagenomics will be the ultimate approach in detecting all micro-organisms (e.g. bacteria, viruses, fungi) in a clinical sample (Hasman et al., 2014). However, analysis of large datasets requires a combination of bioinformatics skills and computational resources that is nowadays mostly absent in diagnostic medical microbiological laboratories. Furthermore, metagenomics approaches are time

consuming as the turnaround time is approximately four to five days.

To fill the gap between the conventional methods (culture and PCR) and metagenomics, a culture-free approach using targeted NGS appears to be an excellent approach to detect and identify bacterial species. Compared to metagenomics, it is faster, less complicated and cheaper, and therefore more likely to get implemented in diagnostic laboratories within a short timeframe. The 16S rRNA gene sequence has been proven to be a reliable genetic marker as it is present in all bacteria, and the function has not changed over time (Patel, 2001). It can be applied directly on clinical materials and has proven to be a valuable supplementary test in daily clinical practice (Schuurman et al., 2004; Srinivasan et al., 2012). However, high sequence similarities in this gene between certain bacterial species do not always lead to an unequivocal identification (Kalia et al., 2016).

Recently, we developed an innovative culture-free 16S-23S rRNA NGS approach for the detection and identification of bacterial species in clinical samples. The method proved to be superior to other commonly used identification methods and correctly identified pathogens in urine samples that were also identified as the cause for urinary tract infections with conventional culture (Sabat et al., 2016). Furthermore, the method allows simultaneous identification of several pathogens in clinical materials that previously would have remained uncultured and PCR negative. Clearly, this will have an enormous clinical impact and will have consequences for patient treatment, including improved antibiotic treatment. Finally, this method will allow clinical microbiology laboratories to implement NGS in their routine diagnostic laboratory and to keep up with technological and bioinformatics developments required to be able to implement metagenomics in diagnostics in the future.

4.5. WGS and taxonomy

In the eighteenth century, Linnaeus (Linnaeus, 1735) provided guidelines for classification of living creatures based on their phenotypic features. A century later, Darwin added the phylogenetic component to the taxonomy (Darwin, 1859). The taxonomy of bacterial species changed dramatically by the introduction of 16S rRNA gene sequencing. Nowadays, WGS is also used to identify and describe new species. By comparing the whole genome sequences of different species with each other, the limited taxonomic resolution of only the 16S rRNA gene can be overcome (Tindall et al., 2010). Another change in taxonomy may be expected when WGS is used for revealing the taxonomy of bacteria.

Indeed, using WGS for taxonomy purposes allows to include more genes to delineate between species than the classical DNA–DNA hybridization or 16S rRNA sequencing methods thereby improving the resolution. Furthermore, as WGS can be used to calculate taxonomic trees based on the whole genome-sequence alignment of all the genes present in the core genome, a more robust tree will be obtained (Daubin et al., 2001). It has already been proposed that descriptions of new taxa should also include a draft genome sequence, with at least 20 times coverage (Thompson et al., 2013).

4.6. Metagenomics in clinical microbiology

As already mentioned, NGS can be applied directly to clinical specimens. Not only by using a targeted NGS approach, but also by sequencing the DNA or RNA from patient samples by shotgun metagenomics sequencing (Fig. 3). Using this method, it is possible to investigate the presence of pathogens and the presence of virulence and/or resistance genes in one sequence run.

A recent study compared the detection of viruses in known respiratory virus-positive samples and not previously analysed

nasopharyngeal swabs by an RNA sequencing-based metagenomics approach with a more conventional molecular method. The data-analyses was performed using Taxonomer, a rapid and interactive, web-based metagenomics data-analyses tool (Flygare et al., 2016). Overall, the metagenomics approach had a high agreement with the molecular method, detected viruses not targeted by the molecular method, and yielded epidemiologically and clinically relevant sequence information (Graf et al., 2016).

Apart from identifying pathogens, a metagenomics approach can also be used to study the resistome. The gut is a known reservoir for antibiotic resistance genes (ARG), and treatment with antibiotics has an impact on the intestinal resistome, which can lead to horizontal gene transfer and the selection of resistant bacteria. A study at the University of Tübingen investigated the presence of ARGs in the gut over a six-day period of ciprofloxacin treatment in two individuals using metagenomics. Furthermore, this study presented a novel method for analysing the determination of antibiotic selection pressure, which can be used in hospitals to compare therapeutic regimens and their effect on the intestinal resistome. This information is important for clinicians to choose antibiotic therapy with a low selective antibiotic pressure on the patient's bacteria in the gut, possibly resulting in a decreased dissemination of antibiotic resistant bacteria (Willmann et al., 2015).

4.7. Determining the transmission of zoonotic micro-organisms from animals to humans

NGS will also reveal more knowledge on zoonotic transmission of micro-organisms. The first studies on this topic were based on low discriminatory methods, such as serotyping (Tenover et al., 1997). More recently, studies using higher discriminatory techniques, such as pulsed-field gel electrophoresis or multi-locus variable number tandem repeat analysis, were used to detect specific bacterial clones in animals and humans (Sabat et al., 2013). However, much remains to be understood, especially when it comes to the frequency of transmission (e.g. single contact or repeated contact with animals or animal products), risk factors associated with the acquisition of a zoonotic microorganism (e.g. risk conducts, such as animal kissing in companion animals, or stool handling in farm animals) and how the use of antibiotics in animals affects the transfer of pathogenic bacteria to humans.

NGS brings a new perspective to these topics. A higher discriminatory power will reveal differences in previously indistinguishable animal and human bacterial strains. This together with epidemiological information allows source tracing of potential zoonotic infections (Harrison et al., 2013). In addition, NGS allows a comprehensive analysis of how antibiotic use manipulates specific microbiota and the consequences for interspecies transmission and will increase the knowledge on microbial evolution through the analysis of bacterial genomes, namely the variable regions, which usually determine host-adaptation and the potential of spread to different hosts (Harrison et al., 2014; Price et al., 2012).

As the patients' safety is depending on its environment, including their contact with food and animals, research projects are currently performed in the UMCG to understand the dynamics of transmission of bacteria between humans, animals and the environment. These studies are performed in collaboration with veterinary research groups and focus on anti-microbial resistant bacteria. In one such study, the *mcr-1* gene was detected by WGS to be present in three *E. coli* strains isolated from retail chicken meat. Although none of the human strains carried this gene, two of the three strains belonged to ST117, a common clone in both poultry and humans, representing a potential public health concern (Kluytmans-van den Bergh et al., 2016a).



For generating NGS data from samples originating from humans, animals, food and the environment, the same laboratory protocol for library preparation can be used, and, after data analyses, information on the presence of specific antibiotic resistance and virulence genes is obtained. Furthermore, NGS makes it possible to standardise typing methods for pathogens (“one test fits all”). The role of NGS in medical microbiology laboratories will increase during the next years, not only for research, but also, and more importantly, for molecular diagnostics, infection prevention, the investigation of outbreaks by the use of a unique outbreak marker approach, the characterization and surveillance of pathogens, the

However, further studies are required to improve the workflow for NGS, in particular shorten the turnaround time for the library preparation and the runs on the NGS platforms, and, at the same time, further reducing costs. Next, automatic pipelines for data-analyses and easy-to-use software for metagenomics have to be developed. Additionally, more established typing schemes for pathogens and cut-off values for these typing schemes have to be established, leading to reference databases with genetic and metadata, and (inter)regional and international collaborations. Importantly, external quality controls for proficiency testing have to be developed. Only then will patient guidance and infection control management at local, (inter)regional and international level,

as well as targeted antibiotic therapy using NGS data become a possibility, leading to personalised microbiology.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding sources

For the writing of this review no specific grant from funding agencies in the public, commercial, or not-for-profit sectors was received.

References

- Aanensen, D.M., Feil, E.J., Holden, M.T., Dordel, J., Yeats, C.A., Fedosejev, A., Goater, R., Castillo-Ramirez, S., Corander, J., Colijn, C., Chlebowicz, M.A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Ahman, J., Matuschek, E., Friedrich, A.W., Parkhill, J., Bentley, S.D., Spratt, B.G., Grundmann, H., European, S.R.L.W.G., 2016. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *MBio* 7, e00444-16.
- Bathoorn, E., Rossen, J.W., Lokate, M., Friedrich, A.W., Hammerum, A.M., 2015. Isolation of an NDM-5-producing ST16 *Klebsiella pneumoniae* from a Dutch patient without travel history abroad, August 2015. *Euro Surveill.* 20, 30040.
- Bertelli, C., Greub, G., 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clin. Microbiol. Infect.* 19, 803–813.
- Buchan, B.W., Ledeboer, N.A., 2014. Emerging technologies for the clinical microbiology laboratory. *Clin. Microbiol. Rev.* 27, 783–822.
- Bush, K., 2013. Proliferation and significance of clinically relevant beta-lactamases. *Ann. N. Y. Acad. Sci.* 1277, 84–90.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., 2005. ACT: the artemis comparison tool. *Bioinformatics* 21, 3422–3423.
- Carver, T., Thomson, N., Bleasby, A., Berriman, M., Parkhill, J., 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119–120.
- Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469.
- de Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., van Schaik, W., Brouwer, E., Rogers, M., Kraat, Y., Bonten, M., Corander, J., Westh, H., Harmsen, D., Willems, R.J., 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. *J. Clin. Microbiol.* 53, 3788–3797.
- Dark, M.J., 2013. Whole-genome sequencing in bacteriology: state of the art. *Infect. Drug Resist.* 6, 115–123.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Races in the Struggle for Life*. John Murray, London.
- Daubin, V., Gouy, M., Perriere, G., 2001. Bacterial molecular phylogeny using supertree approach. *Genome Inform.* 12, 155–164.
- Deurenberg, R.H., Stobberingh, E.E., 2008. The evolution of *Staphylococcus aureus*. *Infect. Genet. Evol.* 8, 747–763.
- ECDC, 2016. Expert Opinion on Whole Genome Sequencing for Public Health Surveillance.
- Edwards, D.J., Holt, K.E., 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb. Inform. Exp.* 3, 2.
- Falgenhauer, L., Waezsada, S.E., Yao, Y., Imirzalioglu, C., Kasbohrer, A., Roesler, U., Michael, G.B., Schwarz, S., Werner, G., Kreienbrock, L., Chakraborty, T., 2016. Colistin resistance gene *mcr-1* in extended-spectrum beta-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. *Lancet Infect. Dis.* 16, 282–283.
- Ferdous, M., Zhou, K., de Boer, R.F., Friedrich, A.W., Kooistra-Smid, A.M., Rossen, J.W., 2015. Comprehensive characterization of *Escherichia coli* O104:H4 isolated from patients in the Netherlands. *Front. Microbiol.* 6, 1348.
- Ferdous, M., Friedrich, A.W., Grundmann, H., de Boer, R.F., Croughs, P.D., Islam, M.A., Kluytmans-van den Bergh, M.F., Kooistra-Smid, A.M., Rossen, J.W., 2016. Molecular characterization and phylogeny of Shiga toxin-producing *Escherichia coli* isolates obtained from two Dutch regions using whole genome sequencing. *Clin. Microbiol. Infect.* 22, 642.e1–642.e9.
- Flygare, S., Simmon, K., Miller, C., Qiao, Y., Kennedy, B., Di Sera, T., Graf, E.H., Tardif, K.D., Kapusta, A., Rynearson, S., Stockmann, C., Queen, K., Tong, S., Voelkerding, K.V., Blaschke, A., Byington, C.L., Jain, S., Pavia, A., Ampofo, K., Eilbeck, K., Marth, G., Yandell, M., Schlager, R., 2016. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* 17, 111.
- Fournier, P.E., Dubourg, G., Raoult, D., 2014. Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Med.* 6, 114.
- Franz, E., Delaquis, P., Morabito, S., Beutin, L., Gobius, K., Rasko, D.A., Bono, J., French, N., Osek, J., Lindstedt, B.A., Muniesa, M., Manning, S., Lejeune, J., Callaway, T., Beatson, S., Eppinger, M., Dallman, T., Forbes, K.J., Aarts, H., Pearl, D.L., Gannon, V.P., Laing, C.R., Strachan, N.J., 2014. Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int. J. Food Microbiol.* 187, 57–72.
- Graf, E.H., Simmon, K.E., Tardif, K.D., Hymas, W., Flygare, S., Eilbeck, K., Yandell, M., Schlager, R., 2016. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *J. Clin. Microbiol.* 54, 1000–1007.
- Harrison, E.M., Paterson, G.K., Holden, M.T., Larsen, J., Stegger, M., Larsen, A.R., Petersen, A., Skov, R.L., Christensen, J.M., Bak Zeuthen, A., Heltberg, O., Harris, S.R., Zadoks, R.N., Parkhill, J., Peacock, S.J., Holmes, M.A., 2013. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol. Med.* 5, 509–515.
- Harrison, E.M., Weinert, L.A., Holden, M.T., Welch, J.J., Wilson, K., Morgan, F.J., Harris, S.R., Loeffler, A., Boag, A.K., Peacock, S.J., Paterson, G.K., Waller, A.S., Parkhill, J., Holmes, M.A., 2014. A shared population of epidemic methicillin-resistant *Staphylococcus aureus* 15 circulates in humans and companion animals. *MBio* 5, e00985–00913.
- Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C.A., Frimodt-Moller, N., Aarestrup, F.M., 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J. Clin. Microbiol.* 52, 139–146.
- Hasman, H., Hammerum, A.M., Hansen, F., Hendriksen, R.S., Olesen, B., Agerso, Y., Zankari, E., Leekitcharoenphon, P., Stegger, M., Kaas, R.S., Cavaco, L.M., Hansen, D.S., Aarestrup, F.M., Skov, R.L., 2015. Detection of *mcr-1* encoding plasmid-mediated colistin-resistant *Escherichia coli* isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro Surveill.* 20, 30085.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P., 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 56, 61–64.
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386.
- Jolley, K.A., Maiden, M.C., 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* 11, 595.
- Junemann, S., Sedlazeck, F.J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., Mellmann, A., Goesmann, A., von Haeseler, A., Stoye, J., Harmsen, D., 2013. Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296.
- Kalia, V.C., Kumar, R., Kumar, P., Koul, S., 2016. A genome-wide profiling strategy as an aid for searching unique identification biomarkers for *Streptococcus*. *Indian J. Microbiol.* 56, 46–58.
- Kluytmans-van den Bergh, M.F., Huizinga, P., Bonten, M.J., Bos, M., De Bruyne, K., Friedrich, A.W., Rossen, J.W., Savelkoul, P.H., Kluytmans, J.A., 2016a. Presence of *mcr-1*-positive *Enterobacteriaceae* in retail chicken meat but not in humans in the Netherlands since 2009. *Euro Surveill.* 21, 30149.
- Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J., Kluytmans, J.A., 2016b. Whole genome multilocus sequence typing of extended-spectrum beta-lactamase-producing *Enterobacteriaceae*. *J. Clin. Microbiol.* 54, 2919–2927.
- Kohl, T.A., Diel, R., Harmsen, D., Rothganger, J., Walter, K.M., Merker, M., Weniger, T., Niemann, S., 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J. Clin. Microbiol.* 52, 2479–2486.
- Laabei, M., Recker, M., Rudkin, J.K., Aldeljawi, M., Gulay, Z., Sloan, T.J., Williams, P., Endres, J.L., Bayles, K.W., Fey, P.D., Yajjala, V.K., Widhelm, T., Hawkins, E., Lewis, K., Parfett, S., Scowen, L., Peacock, S.J., Holden, M., Wilson, D., Read, T.D., van den Elsen, J., Priest, N.K., Feil, E.J., Hurst, L.D., Josefsson, E., Massey, R.C., 2014. Predicting the virulence of MRSA from its genome sequence. *Genome Res.* 24, 839–849.
- Linnaeus, C., 1735. *Systema Naturae, Sive Regna Tria Naturae Systematicae Proposita Per Classes, Ordines, Genera, & Species*. Apud Theodorum Haak, Leiden.
- Liu, W., Li, L., Khan, M.A., Zhu, F., 2012. Popular molecular markers in bacteria. *Mol. Gen. Mikrobiol. Virusol.*, 14–17.
- Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30, 434–439.
- Maiden, M.C., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736.
- Nijhuis, R.H., Oueslati, S., Zhou, K., Bosboom, R.W., Rossen, J.W., Naas, T., 2015. OXY-2-15, a novel variant showing increased ceftazidime hydrolytic activity. *J. Antimicrob. Chemother.* 70, 1429–1433.
- Patel, J.B., 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol. Diagn.* 6, 313–321.
- Price, L.B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P.S., Pearson, T., Waters, A.E., Foster, J.T., Schupp, J., Gillece, J., Driebe, E., Liu, C.M., Springer, B., Zdvoc, I., Battisti, A., Franco, A., Zmudzki, J., Schwarz, S., Butaye, P., Jouy, E., Pomba, C., Porrero, M.C., Ruimy, R., Smith, T.C., Robinson, D.A., Weese, J.S., Arriola, C.S., Yu, F., Laurent, F., Keim, P., Skov, R., Aarestrup, F.M., 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio* 3, e00305–00311.
- Reiss, E., Obayashi, T., Orle, K., Yoshida, M., Zancoppe-Oliveira, R.M., 2000. Non-culture based diagnostic tests for mycotic infections. *Med. Mycol.* 38 (Suppl. 1), 147–159.
- Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H.L., Allerberger, F., Harmsen, D., Mellmann, A., 2015. Defining and evaluating a core genome

- multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *J. Clin. Microbiol.* 53, 2869–2876.
- Sabat, A.J., Budimir, A., Nashev, D., Sa-Leao, R., van Dijk, J., Laurent, F., Grundmann, H., Friedrich, A.W., 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill.* 18, 20380.
- Sabat, A.J., Zanten van, E., Akkerboom, V., Wisselink, G., Slochteren van, K., Boer de, R.F., Friedrich, A.W., Rossen, J.W.A., Kooistra-Smid, A.M.D., 2016. Targeted amplification for bacterial identification at the species-level using next-generation sequencing—increased discrimination of closely related species. ECCMID, E-poster EP0219.
- Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K., Gerstein, M.B., 2011. The real cost of sequencing: higher than you think! *Genome Biol.* 12, 125.
- Schuurman, T., de Boer, R.F., Kooistra-Smid, A.M., van Zwet, A.A., 2004. Prospective study of use of PCR amplification and sequencing of 16S ribosomal DNA from cerebrospinal fluid for diagnosis of bacterial meningitis in a clinical setting. *J. Clin. Microbiol.* 42, 734–740.
- Srinivasan, L., Pisapia, J.M., Shah, S.S., Halpern, C.H., Harris, M.C., 2012. Can broad-range 16S ribosomal ribonucleic acid gene polymerase chain reactions improve the diagnosis of bacterial meningitis? A systematic review and meta-analysis. *Ann. Emerg. Med.* 60, 609–620.
- Tenover, F.C., Arbeit, R.D., Goering, R.V., 1997. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect. Control Hosp. Epidemiol.* 18, 426–439.
- Thompson, C.C., Chimetto, L., Edwards, R.A., Swings, J., Stackebrandt, E., Thompson, F.L., 2013. Microbial genomic taxonomy. *BMC Genom.* 14, 913.
- Tindall, B.J., Rossello-Mora, R., Busse, H.J., Ludwig, W., Kampfer, P., 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266.
- Weterings, V., Zhou, K., Rossen, J.W., van Stenis, D., Thewessen, E., Kluytmans, J., Veenemans, J., 2015. An outbreak of colistin-resistant *Klebsiella pneumoniae* carbapenemase-producing *Klebsiella pneumoniae* in the Netherlands (July–December 2013), with inter-institutional spread. *Eur. J. Clin. Microbiol. Infect. Dis.* 34, 1647–1655.
- Willmann, M., El-Hadidi, M., Huson, D.H., Schutz, M., Weidenmaier, C., Autenrieth, I.B., Peter, S., 2015. Antibiotic selection pressure determination through sequence-based metagenomics. *Antimicrob. Agents Chemother.* 59, 7335–7345.
- Zhou, K., Ferdous, M., de Boer, R.F., Kooistra-Smid, A.M., Grundmann, H., Friedrich, A.W., Rossen, J.W., 2015a. The mosaic genome structure and phylogeny of Shiga toxin-producing *Escherichia coli* O104:H4 is driven by short-term adaptation. *Clin. Microbiol. Infect.* 21 (468), e467–418.
- Zhou, K., Lokate, M., Deurenberg, R.H., Arends, J., Lo-Ten Foe, J., Grundmann, H., Rossen, J.W., Friedrich, A.W., 2015b. Characterization of a CTX-M-15 producing *Klebsiella pneumoniae* outbreak strain assigned to a novel Sequence Type (1427). *Front. Microbiol.* 6, 1250.
- Zhou, K., Lokate, M., Deurenberg, R.H., Tepper, M., Arends, J.P., Raangs, E.G., Lo-Ten-Foe, J., Grundmann, H., Rossen, J.W., Friedrich, A.W., 2016. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum beta-lactamase producing ST15 *Klebsiella pneumoniae*. *Sci. Rep.* 6, 20840.