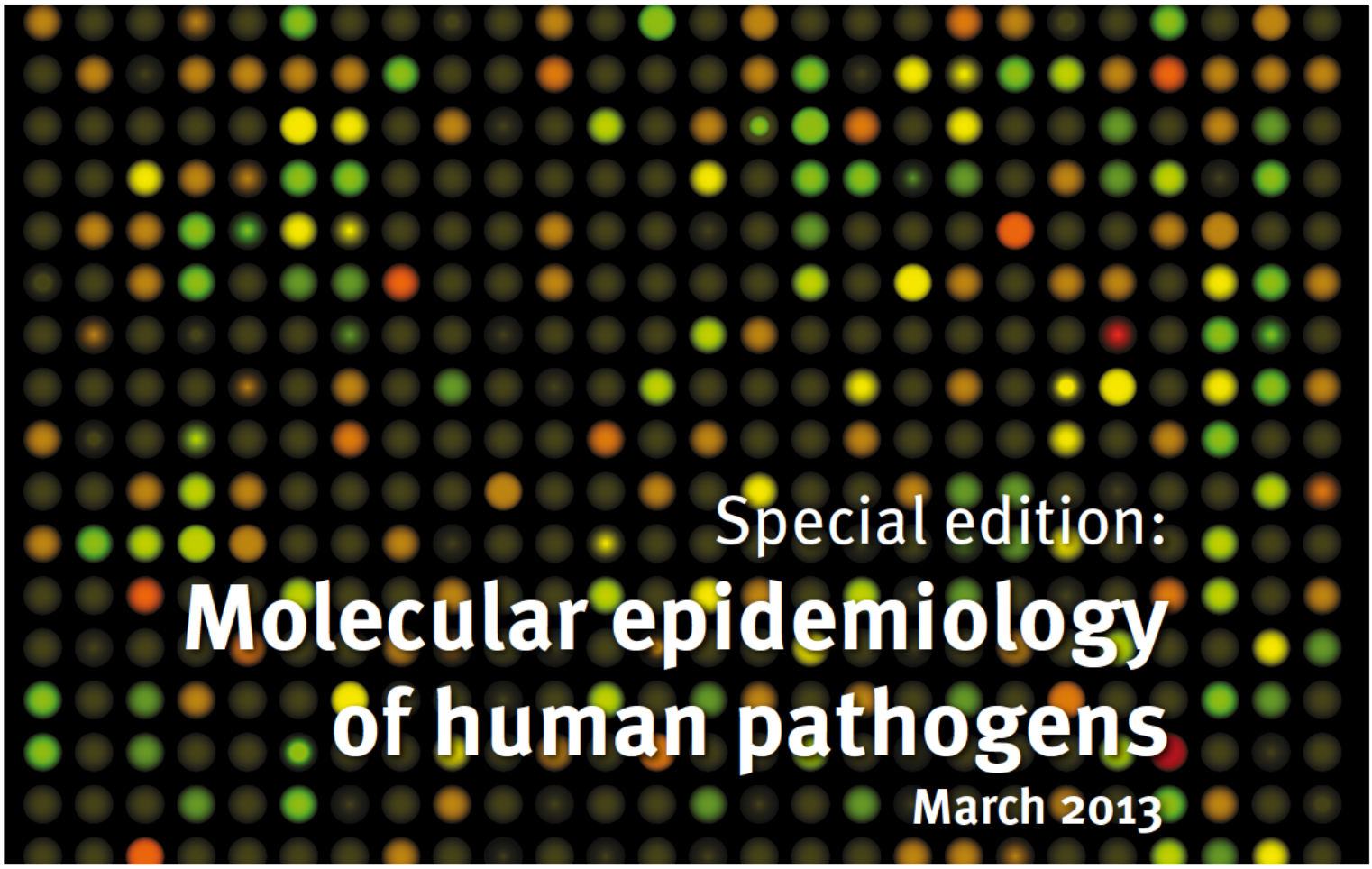
Impact factor 6.15

Eurosurveillance

Europe's journal on infectious disease epidemiology, prevention and control



Special edition:

Molecular epidemiology of human pathogens

March 2013

- This issue presents experiences and examples from various countries in applying molecular methods to a number of diseases in the context of public health as well as several reviews on advantages and challenges of available molecular methods.

www.eurosurveillance.org

Editorial team

Based at the European Centre for
Disease Prevention and Control (ECDC),
171 83 Stockholm, Sweden

Telephone number

+46 (0)8 58 60 11 38 or +46 (0)8 58 60 11 36

Fax number

+46 (0)8 58 60 12 94

E-mail

Eurosurveillance@ecdc.europa.eu

Editor-in-chief

Ines Steffens

Scientific editors

Kathrin Hagmaier

Williamina Wilson

Karen Wilson

Assistant editors

Alina Buzdugan

Ingela Söderlund

Associate editors

Andrea Ammon, Stockholm, Sweden

Tommi Asikainen, Frankfurt, Germany

Mike Catchpole, London, United Kingdom

Denis Coulombier, Stockholm, Sweden

Christian Drosten, Bonn, Germany

Karl Ekdahl, Stockholm, Sweden

Johan Giesecke, Stockholm, Sweden

Herman Goossens, Antwerp, Belgium

David Heymann, London, United Kingdom

Heath Kelly, Melbourne, Australia

Irena Klavs, Ljubljana, Slovenia

Karl Kristinsson, Reykjavik, Iceland

Daniel Lévy-Bruhl, Paris, France

Richard Pebody, London, United Kingdom

Panayotis T. Tassios, Athens, Greece

Hélène Therre, Paris, France

Henriette de Valk, Paris, France

Sylvie van der Werf, Paris, France

Design / Layout

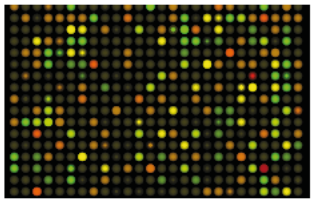
Fabrice Donguy / Arne Haeger

www.eurosurveillance.org

Editorial advisors

Contents

Molecular epidemiology of human pathogens



Microarray

From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases

M J Struelens (Marc.Struelens@ecdc.europa.eu)[1], S Brisse[2]
1. European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
2. Institut Pasteur, Paris, France

Citation style for this article:
Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill. 2013;18(4):pii=20386. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20386

Article published on 24 January 2013

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

As reviewed by Sabat et al., epidemic source tracing requires timely deployment of high resolution typing methods that index variation of genomic elements with a fast molecular clock [1-5]. For outbreak studies, comparative methods, as opposed to library typing methods, are sufficient, and the higher the power to resolve micro-evolutionary distance, the greater the likelihood to decide between alternative transmission hypotheses generated by observational epidemiology [1-6]. Once standardised to enable a uniform genotype nomenclature across laboratories, thereby providing a library typing system, such discriminatory methods can be further applied to control-oriented surveillance [1-5]. Early outbreak detection is achieved by genotyping prospectively as many consecutive cases in a population as possible to identify clusters of clonally linked isolates [5]. Examples include PulseNet, the nationwide food-borne disease surveillance system in the United States [7] as well as national molecular surveillance schemes developed to detect clusters of tuberculosis as described by Fitzgibbon et al. [8]. Library typing systems that use more stable genotypic markers such as bacterial multilocus sequence typing (MLST) are suitable for strategy-oriented molecular surveillance aimed at monitoring secular trends in the evolution of pathogen genotypes and in their distribution over larger geographic and population scales [1-5]. Such molecular surveillance systems can call attention to the emergence of strains with enhanced virulence or drug resistance, help identify risk factors associated with transmission of specific strains, or predict the effectiveness of public health measures such as vaccinations. This approach is well established for global virological surveillance of human and avian influenza. As illustrated by an experience from New-Zealand presented by Muellner et al., a nationwide molecular surveillance of campylobacteriosis using a sequential combination of typing systems can inform both disease control measures and prevention policies by detecting local outbreaks and modelling endemic disease attribution to specific food sources [4]. Structured surveys that combine spatiotemporal mapping of strain genotype and antimicrobial resistance phenotype is a powerful means to monitor the emergence and spread of multidrug-resistant clones across a continent, as reported by Chisolm et al. for *Neisseria gonorrhoeae* in Europe [9].

As summarised by Sabat et al., there have been continuous technological improvements for microbial genomic characterisation in the past decade, moving from fingerprinting methods such as pulsed-field gel electrophoresis of bacterial macrorestriction fragments to more robust, portable and biologically informative assays such as bacterial multilocus variable-number tandem repeat analysis (MLVA) and sequencing of single/multiple loci of both bacterial and viral human pathogens [3-5,9-11]. With the decreasing cost and continuing refinement of high-throughput

genome sequencing technologies, we are now witnessing a quantum leap from genotypic epidemiology to genomic epidemiology as whole viral or bacterial genomes become open to scrutiny at population level. As reviewed by Carrico et al., advances in laboratory typing tools have been enabled by parallel progress in the information technology needed to capture genetic data on pathogens, and in quality control, formatting, storage, management and, most importantly, bioinformatics analysis and real-time electronic data sharing through online databases [10].

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

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

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

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

As opposed to previous sequence-based typing methods, WGS will change the way we look at pathogen diversity in one fundamental way: without an a priori focus on a subset of loci. As all genetic information will be available, it will allow the discovery of novel, unexpected variation, including polymorphisms that evolve during outbreaks or changes that are selected in vivo during infection. Such pathoadaptive changes can result in increased virulence or novel pathophysiological processes. One example of such a micro-evolutionary change is the emergence during influenza A(H1N1)pdm09 epidemic of a quasispecies variant with a haemagglutinin D222G mutation which is associated with modified tissue receptor tropism and severe influenza virus infections, as reported by Rykkvin et al. in this journal [16]. Due to the rapid rate of evolution of viruses and their small genomes, virologists have long been using genome-wide sequencing. The term 'phylodynamics' designates the study of the interplay of epidemiological and evolutionary patterns, pioneered in

virology [17]. Phylodynamics based on WGS of bacterial populations is emerging as a fertile field of investigation for public health microbiology [5-6,14-15].

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

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

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

References

1. Struelens MJ, De Gheldre Y, Deplano A. Comparative and library epidemiological typing systems: outbreak investigations versus surveillance systems. Infect Control Hosp Epidemiol. 1998;19(8):565-9.
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. Goering RV, Köck R, Grundmann H, Werner G, Friedrich AW, on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM). From Theory to Practice: Molecular Strain Typing for the Clinical and Public Health Setting. Euro Surveill. 2013;18(4):pii=20383. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20383
4. Muellner P, Pleydell E, Pirie R, Baker MG, Campbell D, Carter PE, et al. Molecular-based surveillance of campylobacteriosis in New Zealand – from source attribution to genomic epidemiology. Euro Surveill. 2013;18(3):pii=20365. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20365
5. 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
6. Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med. 2012;366(24):2267-75.
7. Swaminathan B, Gerner-Smidt P, Ng LK, Lukinmaa S, Kam KM, Rolando S, et al. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis. 2006;3(1):36-50.
8. Fitzgibbon MM, Gibbons N, Roycroft E, Jackson S, O'Donnell J, O'Flanagan D, et al. A snapshot of genetic lineages of Mycobacterium tuberculosis in Ireland over a two-year period, 2010 and 2011. Euro Surveill. 2013;18(3):pii=20367. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20367
9. Chisholm SA, Unemo M, Quaye N, Johansson E, Cole MJ, Ison CA, et al. Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrug-resistant clone. Euro Surveill. 2013;18(3):pii=20358. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20358
10. Carriço JA, Sabat AJ, Friedrich AW, Ramirez M, on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM). Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. Euro Surveill. 2013;18(4):pii=20382. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20382
11. Lindstedt BA, Torpdahl M, Vergnaud G, Le Hello S, Weill FX, Tietze E, Malorny B, Prendergast DM, Ní Ghallchóir E, Lista RF, Schouls LM, Söderlund R, Börjesson S, Åkerström S. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill. 2013;18(4):pii=20385. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20385
12. Jolley KA, Maiden MC. Automated extraction of typing information for bacterial pathogens from whole genome sequence data: Neisseria meningitidis as an exemplar. Euro Surveill. 2013;18(4):pii=20379. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20379
13. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog. 2012;8(6):e1002776.
14. Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. Current application and future perspectives of molecular typing methods to study Clostridium difficile infections. Euro Surveill. 2013;18(4):pii=20381. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20381
15. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet. 2012;45(1):109-13.
16. Rykkvin R, Kilander A, Dudman SG, Hungnes O. Within-patient emergence of the influenza A(H1N1)pdm09 HA1 222G variant and clear association with severe disease, Norway. Euro Surveill. 2013;18(3):pii=20369. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20369
17. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303(5656):327-32.
18. 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.
19. Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics. 2012;13:338.
20. Niesters HG, Rossen JW, van der Avoort H, Baas D, Benschop K, Claas EC, Kroneman A, van Maarseveen N, Pas S, van Pelt W, Rahamat-Langendoen JC, Schuurman R, Vennema H, Verhoef L, Wolthers K, Koopmans M. Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories . Euro Surveill. 2013;18(4):pii=20387. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20387
21. Rump B, Cornelis C, Woonink F, Verweij M. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. Euro Surveill. 2013;18(4):pii=20384. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20384
22. Palm D, Johansson K, Ozin A, Friedrich AW, Grundmann H, Larsson JT, et al. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011. Euro Surveill. 2012;17(2):pii=20054. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20054
23. van Walle I. ECDC starts pilot phase for collection of molecular typing data. Euro Surveill. 2013;18(3):pii=20357. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20357
24. Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, et al. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg Infect Dis. 2012;18(11):e1

Overview of molecular typing methods for outbreak detection and epidemiological surveillance

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

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 Quimica e Biologica, 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 Overdevest 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 approximatively 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) 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 *fljB* 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.

The great advantage of MLST based on seven housekeeping 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/bigsdb/), 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 interlaboratory 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/). 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, bioinformatic 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 Heath-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. FAFLP: 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. C*ompa*rison 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. Overdevest 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-Verheuvel 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 meticillin-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. Infec 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 noncapsulated Haemophilus influenzae and determination of phylogenetic relationships by multilocus sequence typing. J Clin Microbiol. 2003;41(4):1623-36.
65. Salcedo C, Arreaza L, Alcala 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, Kornschober 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-Smidt 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, Ehricht 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, Potamousis KD, 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.

Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution

J A Carriço (jcarrico@fm.ul.pt)[1], A J Sabat[2], A W Friedrich[2], M Ramirez[1], on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM)[3]

1. Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal
2. Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
3. European Society for Clinical Microbiology and Infectious Diseases, Basel, Switzerland

Citation style for this article:
Carriço JA, Sabat AJ, Friedrich AW, Ramirez M, on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM). Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. Euro Surveill. 2013;18(4):pii=20382. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20382

Article submitted on 29 June 2012 / published on 24 January 2013

Advances in typing methodologies have been the driving force in the field of molecular epidemiology of pathogens. The development of molecular methodologies, and more recently of DNA sequencing methods to complement and improve phenotypic identification methods, was accompanied by the generation of large amounts of data and the need to develop ways of storing and analysing them. Simultaneously, advances in computing allowed the development of specialised algorithms for image analysis, data sharing and integration, and for mining the ever larger amounts of accumulated data. In this review, we will discuss how bioinformatics accompanied the changes in bacterial molecular epidemiology. We will discuss the benefits for public health of specialised online typing databases and algorithms allowing for real-time data analysis and visualisation. The impact of the new and disruptive next-generation sequencing methodologies will be evaluated, and we will look ahead into these novel challenges.

Introduction

In the past twenty years, the advances in several fields of biology, molecular biology in particular, led to an increased capacity to generate data. This resulted in the accumulation of large datasets and the need to store, manage and analyse them. This was the starting point for the development of the multidisciplinary field of bioinformatics. Hesper and Hogeweg originally coined the term bioinformatics in 1970 [1]. It was broadly defined as "the study of informatics processes in biotic systems". But it was the convergence of mathematicians, computer scientists, physicists, biologists, chemists and health professionals for the analysis of the biological data generated in the genomic revolution that resulted in the diverse disciplines comprised within bioinformatics. The field can also be subdivided into two large, interrelated subareas: data management, encompassing the creation and management of databases for biological data, and data analysis, ranging from the creation of mathematical and statistical models to computational tools and data mining techniques.

In bacterial molecular epidemiology, bioinformatics drove the creation of online databases for microbial typing data (e.g. antibiotic resistance profiles, phage typing, serotyping or other phenotypic information), the analytic methodologies for gel-based molecular typing techniques and the study and analysis of phylogenetic inference models.

In this review we aim to provide a perspective on the bioinformatics tools that have been applied in the field of bacterial molecular epidemiology. We will explore their applications in public health, documenting how they have changed and discussing possible avenues for future research and development in the field.

Online databases for bacterial typing

Microbial typing methods allow the characterisation of bacteria to the strain level, providing researchers with important information for surveillance of infectious diseases, outbreak investigation and control. These methods offer insights into the pathogenesis and natural history of an infection, and into bacterial population genetics [2,3], areas of research that have an important impact on human health issues such as the development of vaccines or novel antimicrobial drugs [4], with significant social and economical implications.

Molecular typing methods, such as pulsed-field gel electrophoresis (PFGE), provided the intra- and inter-laboratory reproducibility needed to create databases of isolates that could be used for longitudinal studies [3]. This allowed for bacterial typing to extend beyond outbreak investigation. Results were originally stored in local databases, using specialised software

Table 1

Online molecular typing databases

Method	Database	URL
MLST	MLST.net	http://www.mlst.net
	Pubmlst.org	http://www.pubmlst.org
	Institut Pasteur MLST	http://www.pasteur.fr/mlst/
	European Working Group for Legionella Infections Sequence-based typing database	http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php
	Environmental Research Institute, University College Cork	http://mlst.ucc.ie/
MLVA	MLVAbank	http://minisatellites.u-psud.fr/MLVAnet/
	Groupe d'Etudes en Biologie Prospective	http://www.mlva.eu
	MLVAplus	http://www.mlvaplus.net/
	Institute Pasteur MLVA: MLVA-NET	http://www.pasteur.fr/mlva
	MLVA.net	http://www.mlva.net
ccrB typing	Staphylococci ccrB sequence typing	http://www.ccrbtyping.net/
dru typing	dru typing database	http://www.dru-typing.org
spa typing	Ridom Spa Server	http://spaserver.ridom.de/
CRISPR typing	CRISPRdb	http://crispr.u-psud.fr/crispr/

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis.

such as BioImage Whole Band Analyzer (Genomic Solutions, Inc, Ann Arbor, MI, currently discontinued) and GelCompar (currently GelComparII or Bionumerics from Applied Maths, Ghent, Belgium). These pieces of software, which integrated rudimentary database management and gel image analysis, were in fact the first widely adopted bioinformatics tools used in the field.

The ability to share information using the Internet led to the next step: the evolution of those software applications to distributed systems in which nationwide or worldwide comparisons could be performed. PulseNet, the molecular subtyping network for foodborne bacterial disease [5] was the first network that created local and central databases where laboratories from across the United States (US), could securely query nationwide data and compare their local samples. PulseNet is a governmental network initiated by the US Centers for Disease Control and Prevention and laboratories in several state health departments in the US, but has evolved to PulseNet International (www.pulsenetinternational.org/) [6]. PulseNet was created based on standardised PFGE protocols for the identification of pathogenic food-borne bacteria, relying on specifically trained technical personnel, but nowadays also integrates information obtained by other typing methods.

The network derives its strength from a series of bioinformatics techniques, implemented in the Bionumerics software, that range from optimised algorithms for gel image analysis and comparison to database management and secure sharing of data. The PulseNet online information system became the first distributed database for microbial typing with a direct application in public health and remains an example of the successful application of bioinformatics in typing and molecular epidemiology.

What the PulseNet distributed network achieved for PFGE, was much more simply achieved for multilocus sequence typing (MLST) [7], due to the inherent portability of sequence data (i.e. data easily transferrable between different systems). MLST is based on the analysis of allelic profiles generated by comparing sequences to an online repository. In contrast to PulseNet, MLST websites host publicly accessible databases where any laboratory can submit data, while PulseNet is only accessible by their member laboratories due to privacy and confidentiality issues (Table 1).

The ability to easily share sequence data through the Internet [8,9] is one of the main characteristics that made MLST the method of choice for clonal identification and tracking for many bacterial species. Currently available MLST databases (Table 1) are more commonly used for nomenclature purposes and may not reflect clonal abundance. The portability that is characteristic of MLST allows disambiguation when analysing and comparing results. Another important feature that contributed to its success was the possibility to infer patterns of phylogenetic descent through comparison of the allelic profiles. Even though MLST became the gold standard for long-term epidemiological surveillance of several species, PFGE remains important for outbreak

detection because it often has higher discriminatory power.

One example of an MLST online database, with proven use in public health, is the European Working Group for Legionella Infections (EWGLI) database, currently part of the European Legionnaire's Disease Surveillance Network (ELDSnet). This typing scheme and database successfully identified sources of infection, by determining clonal identity between environmental and patient isolates of *Legionella pneumophila* [10].

Several other sequence-based typing methodologies with online databases have become available. In contrast to MLST, the majority of these methods are only available for certain species, since they focus on non-housekeeping genes, and most are single locus sequence typing (SLST) schemes.

Taking *Staphylococcus aureus* as an example, several SLST were developed in the past decade. Two methods based on variable-number of tandem repeats (VNTR) were proposed, one relying on the direct repeat unit (*dru*) VNTR region adjacent to IS*431* in SCC*mec* [11], and the other based on the analysis of repeat patterns in the *spa* gene, the now widely used *spa* typing [12]. A major factor for the widespread use of *spa* typing was the implementation of a user-friendly software, Ridom StaphType. This tool allows the automatic assignment of a *spa* type from a DNA sequence in Fasta format or directly from chromatograms, through comparison with the centralised SpaServer [13]. Another SLST is *ccrB* typing [14], originally developed for meticillin-resistant *S. aureus* (MRSA), but extended and applicable to all staphylococci containing the *mecA* gene, the determinant of meticillin resistance. Also this method benefits from online databases and tools.

A multilocus methodology that has recently shown promise for several bacterial species is multilocus VNTR analysis (MLVA). Similarly to MLST it produces a numeric profile, in this case of the number of repeats at each locus that can unambiguously identify a given strain (MLVA type). Its appeal derives from providing a highly discriminatory method that shows high congruence with MLST results for several bacterial species [15], but is less expensive since sequencing of the loci is not necessary. Databases for a variety of schemes and bacterial species have been made available by several institutions (Table 1). Some of these online databases offer users the possibility to create their own private or public database like MLVAbank [16], MLVAplus or MLVA-NET [17]. A particular application of an MLVA scheme is the MIRU-VNTRplus Internet application for *Mycobacterium tuberculosis* [18,19].

Recently, a new sequence-based typing methodology was proposed using clustered regularly interspaced short palindromic repeats (CRISPR), a specific family of DNA repeats, conferring resistance to foreign DNA such as plasmids and phages. A database and tools are also available online (Table 1).

With next generation sequencing (NGS) technologies comes the ability to quickly obtain complete or nearly complete genome sequences of thousands of individual strains. In spite of the great promise of these approaches, it is still unclear how whole-genome data on bacterial pathogens will be shared and used for bacterial population surveillance and possible applications in public health.

BIGSdb, is a database system recently proposed to handle NGS data of microbial genomes and perform analyses focused on extended MLST typing approaches, which can comprise thousands of genes, and also on other population analysis methodologies [20]. One such scheme is ribosomal MLST [21] that, by focusing on the same ribosomal genes, allows a universal characterisation of bacteria, encompassing all levels of bacterial diversity, from domain to strain.

In highly monomorphic and slowly evolving bacterial species such as *M. tuberculosis* or Bacillus anthracis, identification of single nucleotide polymorphisms (SNPs) by comparison to a defined archetypal strain, could also be a basis for analysis, imposing different requirements on an online database.

Tools for data analysis

The cornerstone of molecular epidemiology is the ability to compare the classification results obtained by a given typing method for two or more distinct isolates and to measure their relatedness. With that information, one can then support an epidemiological investigation or raise a hypothesis about phylogenetic relationship. In this section we will describe several of the techniques developed in the last decades and used in the analysis of molecular typing data.

The first methodologies used in analysis of the phenotypic and genotypic data, were classical techniques used in numerical taxonomy [22], a field pioneered by P. Sneath and R. Sokal. The most popular are hierarchical clustering methods, which result in a unique tree representing the relationships between isolates, commonly called dendrogram or phenogram. From that tree, groups of related isolates are defined by a similarity level cut-off. These are mathematical methods that were implemented in generic statistical software or custom-made computer programmes. However, for the analysis of gel-based typing data, an integrated solution of image analysis and normalisation was needed prior to data analysis. This led to the development of the first tools specific for the analysis of gel-based typing methods. They allowed the quantitative analysis of large numbers of isolates and their comparison with databases of already characterised strains for gel-based methodologies such as PFGE, random amplification of polymorphic DNA (RAPD) [23], amplified fragment length polymorphism (AFLP) [24] or any

Table 2

Currently available software for the analysis of typing results

Application	Software	URL	Availability
Gel analysis	GelCompare II	http://www.applied-maths.com/gelcompar-ii	Commercial
	Phoretix 1D	http://www.totallab.com/products/1d/	
	Gel-Pro Analyzer 4.5	http://www.mediacy.com/index.aspx?page=GelPro	
Sequence assembly and analysis	Lasergene	http://dnastar.com	
	CLCbio workbench	http://www.clcbio.com/products/clc-main-workbench/	
	Geneious	http://www.geneious.com/	
Multiple	Bionumerics	http://www.applied-maths.com/bionumerics	
Phylogenetic inference	eBURST v3	http://eburst.mlst.net	Freeware
	MEGA 5	http://megasoftware.net/	
	PHYLOViZ 1.0	http://www.phyloviz.net	
	Structure 2.3.3	http://pritch.bsd.uchicago.edu/structure.html	
	BAPS 5.4	http://www.helsinki.fi/bsg/software/BAPS/	
	ClonalFrame 1.2	http://www.xavierdidelot.xtreemhost.com/clonalframe.htm	
Typing methods comparison	Ridom Epicompar	http://www.ridom.de/epicompare/	
	Comparing Partitions	http://www.comparingpartitions.info	
Recombination assessment	RDP3	http://darwin.uvigo.es/rdp/rdp.html	
Sequence comparison and analysis	Mauve	http://gel.ahabs.wisc.edu/mauve	

restriction fragment length polymorphism (RFLP) methodology. Presently, the most widely used and complete software solution for the analysis of gel-based typing methods is the commercially available Bionumerics, as it incorporates several hierarchical clustering algorithms for the analysis of typing data, as well as algorithms for the analysis of DNA sequences (Table 2).

With the appearance of MLST, new analysis methodologies were developed that tried to incorporate a model of bacterial evolution and spread. eBURST (based upon related sequence types) [25] implements a simple model for the emergence of clonal complexes [26,27]: a given genotype increases in frequency in the population and becomes a founder clone, and this increase is accompanied by a gradual diversification of that genotype, by mutation or recombination, forming a cluster of phylogenetically related strains. Software that performs eBURST analysis is available as freeware (Table 2).

The eBURST algorithm was further extended by goeBURST [28], a global optimal implementation of the eBURST algorithm that guarantees a unique solution for the BURST rules, while simultaneously allowing an assessment of the validity of each drawn link. The goeBURST algorithm is not exclusive for the analysis of MLST sequence types (ST) and can also be used in the analysis of any other sequence-based typing method that produces an allelic profile, such as MLVA or even SNP data from NGS methods. goeBURST also clarified the relationship between BURST rules and the use of minimum spanning trees (MSTs), another commonly used method in the analysis of sequence-based typing methods. It showed that the definition of clonal complexes by goeBURST is identical to pruning an MST at a chosen number of differences in the profiles that are being compared. That MSTs are easy to interpret has made them one of the preferred representation methods of the relationships inferred from SNP data in a variety of studies [29-31]

Although eBURST or goeBURST have been used extensively and successfully for determining the genetic population structure of many bacterial species, they also have limitations. As with other methods of phylogenetic reconstruction, the BURST rules do not specifically take into account recombination. Recombination is increasingly recognised as a major force in bacterial evolution, and when it involves segments of DNA larger than the internal gene fragments analysed by MLST, this will lead to the presence of the same alleles in strains from different genetic lineages. Horizontal gene transfer can therefore result in STs that have similar allelic profiles due to recombination, rather than recent shared ancestry. This is particularly true for some bacterial species such as Enterococcus faecium and Burkholderia pseudomallei in which recombination occurs with very high rates [32]. In other instances, recombination was even shown to occur between different species of the same genus [33]. To highlight recombination occurring within the analysed fragments different methods can be used, many are implemented in the software RDP3 [34], while traditional phylogenetic methods are helpful in detecting recombination between different species. An important set of tools are implemented in the software MEGA (Molecular Evolutionary Genetics Analysis) [35].

For the analysis of *spa* typing data, an algorithm was proposed to create clonal complexes from the sequence of repeats, based on an evolutionary model of repeated excision and duplication as well as single nucleotide substitutions and indels (insertions or deletions) (EDSI) [36]. This approach is available in the BURP (based upon repeat pattern) algorithm [37], implemented in the Ridom StaphType software, but could also be applied to other VNTR analysis.

An important aspect in the analysis of typing data is the integration of the algorithm results with epidemiological data. This is usually done by annotation of the resulting trees or dendrograms. Bionumerics offers that possibility in its multiple analysis algorithms. The freely available PHYLOViZ software [38] offers a more dynamic interface for the integration of this information into a goeBURST analysis, in the expansion of the goeBURST rules to any number of loci and in MSTs.

In epidemiological studies, the spatial component is of great importance. The ability to monitor the geographic spread of clones at different levels (cities, countries, continents or worldwide) can provide a perspective of the dissemination of successful clones. The website www.spatialepidemiology.net provides users with a map-based interface that allows the display and analysis of epidemiological data for infectious diseases. It was used by the European Antimicrobial Resistance Surveillance System (EARSS) [39] to provide a genetic snapshot of the *S. aureus* population causing invasive disease in Europe, plotting *spa* typing data, antibiotic resistance and other epidemiologically relevant data [40]. The website can also be connected to the EpiCollect system [41], allowing the real-time collection and annotation of data using any browser or smartphone.

The growing availability of sequence data also led to the increased popularity of model-based statistical analysis approaches. These focus on the use of Bayesian theory to infer the most probable population structure. The software applications STRUCTURE [42], Clonalframe [43] and Bayesian Analysis of Population Structure (BAPS) [44,45] are freely available, but have high computational requirements for large datasets. STRUCTURE and BAPS were initially proposed for classical population genetic analysis and try to infer possible population structures by identifying admixture events in the population history. Clonalframe was proposed for the analysis of MLST sequence data or alignments of multiple bacterial genomes and takes into account the possibility of recombination between sequences. More recently, BAPS was also adapted to detect and represent recombination between different populations and subpopulations [46] using MLST sequences as input. These methodologies can provide a much finer picture of how the phenomena shaping population structure interact and how they influence the final population [47-49], but the computational needs and complex analysis of results still limit their application in the field of bacterial epidemiology.

Not all bioinformatics tools in molecular epidemiology were initially designed for clonal inference from typing data. Two freely available tools were developed with the goal of providing a quantitative comparison of typing methods. Ridom Epicompare is a stand-alone software that allows the calculation of Simpson's index of diversity [50] and 95% confidence intervals [51] for a typing method, and the concordance indexes of Rand [52], adjusted Rand [53] and Wallace [54] for the assessment of congruence between typing methods [55]. The website www.comparingpartitions.info extends the features of Epicompare, by implementing confidence intervals for Wallace [56] and adjusted Rand [57] indices, as well as an adjusted Wallace coefficient and respective 95% confidence intervals [58]. These discriminatory and concordance indexes are now being used for evaluating the adequacy of a method for epidemiological typing. More recently these indexes were used to evaluate cut-off criteria for defining groups. This was done for multilocus variable-number tandem repeat fingerprinting (MLVF) patterns for *S. aureus* typing, including analyses of outbreaks and strain transmission events [59] as well as for PFGE [60], and also for defining clones in *Staphylococcus epidermidis* [61].

Bioinformatics for molecular epidemiology: the way forward

The advances in the last two decades in DNA sequencing capacity and bioinformatics led to an increase in the number of databases and software tools for microbial typing methods. The ability to freely share sequence data over the Internet, pioneered by MLST databases, was the turning point for the definition of a common language for the identification of bacterial clones.

However, the currently available databases suffer from several drawbacks. In some cases, data submission and curation protocols still rely heavily on human input with the exchange of files by email or other non-automated processes that are prone to human error and lead to extended response times by curators. Another missing feature is the absence of application programming interfaces for automatic querying and of standardised data sharing formats. These limitations make data collation a difficult and laborious manual process that requires integrating data from different databases and preparing them for analysis by available software. Consequently, a wealth of data is left largely inaccessible and unexplored.

The first step in tackling these problems is the definition of a common language to exchange data between databases and between databases and software. This is the starting point for the creation of database interoperability, i.e. the ability of tools in one database to query another, allowing for transparent data integration.

Current concepts and technologies for data integration are focused in the Semantic Web [62] and Linked Open Data concepts [63]. These concepts envision a data-centric approach with loosely standardised formats for information exchange, based on explicit data descriptions [64]. To achieve these goals, an ontology of terms in the field must be explicitly described. Ontologies provide a formal, standardised representation of the data and the relationships between the data entities [65]. Recently, the prototype of an ontology for microbial typing was proposed and made publicly available at www.phyloviz.net/typon/ [66]. The use of the ontology and the concepts of Linked Data for the construction of webservices for data exchange and validation could prove fundamental for the integration of the present techniques with the new NGS methods. This would allow NGS databases and data analysis algorithms to be validated against the large body of data available in existing databases.

The potential of NGS technology to become the ultimate methodology for bacterial identification and typing has been recognised by the scientific community, and the first steps towards its application have been taken.

NGS data result from a plethora of different technologies, each with its own strengths and caveats [67]. Running a single NGS analysis of an isolate will generate an amount of data that is orders of magnitude greater than that generated by other typing methods. As an example, the reads of a single bacterial genome with 100-fold coverage, will occupy around 200 MB of disk space. To handle this amount of data requires a complex IT infrastructure that was not necessary before. This also generates computational challenges that must be addressed by specialised software. Cloud computing and the use of high performance computing facilities will mitigate this problem, but are not a substitute for optimised algorithms. Stimulating collaborations between computer scientists and mathematicians with interest in biological problems, and developing specific training programmes will be key to attaining this goal.

Since the technology has been in constant evolution and the algorithms are evolving with it, there is currently no stable pipeline for the analysis of NGS data [68]. Due to limited availability of expertise in this area, centralised hubs for NGS application in diagnosis and public health have been proposed [69]. As the technology matures, the situation may change, allowing the deployment of NGS at hospital level. Recent releases of commercial Windows-based software with a menu-driven approach are a first step towards this goal (Table 2). However, it is important to note that at the current pace of innovation in NGS, these platforms frequently incorporate already superseded versions of algorithms that are under constant development in UNIX-based counterparts, less user-friendly, but freely available.

There are already several successful applications of NGS to a variety of public health problems, ranging from outbreak or short-term epidemiology investigations, to the discovery of unsuspected zoonosis cases and long-term epidemiology studies.

An event that received considerable media coverage was the outbreak of Escherichia coli O:104 haemolytic-uraemic syndrome in Germany that started in May 2011. Due to the pioneering crowdsourcing efforts in annotating an early released genome of an outbreak isolate and subsequent follow-up analyses [70,71], it was possible to promptly develop a diagnostic PCR to identify outbreak isolates. Subsequent studies were able to propose that the outbreak strain, E. coli O104:H4, had emerged due to horizontal gene exchange, shedding novel light on the emergence of new pathogens [72].

A recent pilot study focusing on the nosocomial pathogens MRSA and Clostridium difficile evaluated the feasibility of using benchtop sequencers for outbreak detection and surveillance at hospital level [73]. The ability to further discriminate isolates grouped together by other typing methods allowed a better understanding of the chains of transmission and supported infection control measures. Similar results were achieved when tracing an MRSA outbreak in a neonatal ward [74].

Long-term epidemiological studies have also benefited from NGS technology. The evolution of extremely successful and clones with worldwide dissemination has been followed for MRSA and *Streptococcus pneumoniae* [75,76]. Using SNP to identify phylogenetic relationships, these studies mapped the acquisition of mobile genetic elements and the fast-paced evolution of surface antigens that had frequently confounded previous analyses.

Most intriguing was the use of NGS to identify a probable zoonotic origin for autochthonous leprosy cases in the southern United States [77]. The study identified a unique genotype in this geographic area that also occurred in the armadillo population, strongly suggesting a zoonotic origin and a potential avenue for the control of this infection.

Two recent international meetings discussed and defined roadmaps in bacterial genomic identification and outbreak detection through the use of NGS.

The National Food Institute at the Technical University of Denmark issued a consensus report from an expert meeting on the perspectives of a global, real-time microbiological genomic identification system [78]. In this report it is recognised that within 5 to 10 years, DNA sequencers will likely be a common tool in clinical microbiology laboratories, and that the limiting factor will not be the cost of whole genome sequencing, but the creation of standardised pipelines to handle

the large amounts of data generated. It was also highlighted that a clear and widely accepted concept of the term 'clone' was needed, and that the comparison with data from existing databases (for example MLST) will play a crucial part in validating whole-genome approaches and providing the link with currently accepted and validated methodologies. It was further recognised that achieving this goal required "a global system or at least inter-operable systems to aggregate, share, mine and translate the genomic data to direct part of the genomics efforts to address global public health and clinical challenges, a high impact area in need of focused effort" [78].

A follow-up meeting was held in Washington, under the auspices of the United States Food and Drug Administration, also with the objective of establishing consensus guidelines in the field, focusing on NGS technology for outbreak detection. One of the most debated topics was the future development of databases for NGS data. The need for publicly available data repositories with NGS data from all bacterial domains was reinforced as a prerequisite for the development of new analysis methods.

These needs were also recently recognised in an expert consultation on molecular epidemiology hosted and organised by the European Centre for Disease Prevention and Control (ECDC) [79].

As more data becomes available, it is clear that molecular epidemiology will also benefit from closer integration with basic research in evolution and population biology. Changes in databases and analysis tools will be needed to bring about this integration in order to empower stakeholders in everyday public health decisions.

Tools are being developed to integrate different sources of molecular epidemiology data as well as other metadata (place, time, etc). However, these efforts are still in their infancy, and greater emphasis will need to be placed on the integration of different information sources in the analysis algorithms. Through the combined analysis of this information we can obtain knowledge of the epidemiology of infectious diseases. In particular, the broader use of geographic information in phylo-geographical approaches will allow a better understanding of the spread of particular clones [80].

Conclusions

Epidemiology has come a long way since John Snow investigated the cholera epidemic in Soho, London, in 1854. From hand-plotting cases on a map, we have come to depend on computing power and complex bioinformatics algorithms to make sense of the wealth of available molecular epidemiology data. It is clear that bioinformatics tools have raised the public health impact of the widely used typing methods. Similarly, the NGS revolution will not be extensively available to health professionals until several bioinformatics challenges have been solved and the results can be reported in a way that can be acted upon in everyday practice.

Integration of data of already established microbial typing methods, genomic and epidemiological databases and NGS data will be the next frontier in bacterial epidemiology. Once NGS becomes widely adopted, the development of software that analyses information from different data sources will be key to the synthesis of available knowledge. The public health community must also define standards for analysis and reporting, in order to produce the desired reproducibility and common language needed for typing based on NGS to be useful in clinical settings. More than ever, the need for a convergence of specialists of numerous disciplines in the field of bioinformatics will be fundamental to solve these challenges.

References

1. Hesper B, Hogeweg P. Bioinformatica: een werkconcept [Bioinformatics: a working concept]. Kameleon 1970; 1(6):28–9. Dutch.
2. van Belkum A, Struelens M, de Visser A, Verbrugh H, Tibayrenc M. 2001. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin Microbiol Rev. 2001;14(3):547–60.
3. Struelens M. 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect 1996;2(1):2-11.
4. McKnew DL, Lynn F, Zenilman JM, Bash MC. Porin variation among clinical isolates of Neisseria gonorrhoeae over a 10-year period, as determined by Por variable region typing. J Infect Dis. 2003;187(8):1213–22.
5. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV, CDC PulseNet Task Force. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg Infect Dis 2001;7(3):382–9.
6. Swaminathan B, Gerner-Smidt P, Ng L-K, Lukinmaa S, Kam K-M, Rolando S, Gutiérrez EP, Binsztein N. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis. 2006;3(1):36–50.
7. Maiden M, Bygraves J, Feil EJ, Morelli G, Russell J, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA. 1998;95(6):3140–5.
8. Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol. 1999;2(3):312–6.
9. Maiden MC.. Multilocus sequence typing of bacteria. Annu Rev Microbiol. 2006;60:561–88.
10. Allerberger F. Molecular typing in public health laboratories: from an academic indulgence to an infection control imperative. J Prev Med Public Health. 2012;45(1):1–7.
11. Goering R, Morrison D, Doori Al Z, Edwards G, Gemmell C. Usefulness of mec-associated direct repeat unit (dru) typing in the epidemiological analysis of highly clonal methicillin-resistant Staphylococcus aureus in Scotland. Clin Microbiol Infect. 2008;14(10):964–9.
12. Frénay HM, Bunschoten AE, Schouls LM, van Leeuwen WJ, Vandenbroucke-Grauls CM, 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.
13. 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.
14. Oliveira DC, Santos M, Milheiriço C, Carriço JA, Vinga S, Oliveira AL, et al.. CcrB typing tool: an online resource for staphylococci ccrB sequence typing. J Antimicrob Chemother 2008;61(4):959–60.
15. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, van Santen-Verheuvel 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.
16. Grissa I, Bouchon P, Pourcel C, Vergnaud G. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. Biochimie. 2008;90(4):660–8.
17. Guigon G, Cheval J, Cahuzac R, Brisse S. MLVA-NET--a standardised web database for bacterial genotyping and surveillance. Euro Surveill. 2008;13(19) pii=18863. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18863
18. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol. 2008;46(8):2692–9.
19. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res. 2010;38(Web Server issue):W326–31.
20. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.
21. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 2012;158(pt 4):1005–15.
22. Sneath PHA, Sokal RR. Numerical Taxonomy: The principles and practice of numerical classification. 1st ed. San Francisco:W.H. Freeman & Co;1973.
23. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 1990;18(22):6531–5.
24. Vos P, Hogers R, Bleeker M, Reijans M, van De Lee T, Hornes M, et al. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23(21):4407–14.
25. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004;186(5):1518–30.
26. Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NP, Enright MC, et al. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc Natl Acad Sci U S A. 2001;98(1):182–7.
27. Smith JM, Feil EJ, Smith NH. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays. 2000;22(12):1115–22.
28. Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. 2009;10:152.
29. Nübel U, Roumagnac P, Feldkamp M, Song J, Ko K, Huang Y, et al. Frequent emergence and limited geographic dispersal of methicillin-resistant Staphylococcus aureus. Proc Natl Acad Sci U S A. 2008;105(37):14130-5.
30. Roumagnac P, Weill F-X, Dolecek C, Baker S, Brisse S, Chinh NT, et al. Evolutionary history of Salmonella typhi. Science. 2006;314(5803):1301–4.
31. Baker S, Holt K, van de Vosse E, Roumagnac P, Whitehead S, King E, et al. High-throughput genotyping of Salmonella enterica serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban District of Jakarta, Indonesia. J Clin Microbiol. 2008;46(5):1741–6.
32. Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol. 2007;7:30.
33. McMillan DJ, Bessen DE, Pinho M, Ford C, Hall GS, Melo-Cristino J, et al. Population genetics of Streptococcus dysgalactiae subspecies equisimilis reveals widely dispersed clones and extensive recombination. PLoS One. 2010;5(7):e11741.
34. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26(19):2462–3.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.
36. Sammeth M, Stoye J. Comparing tandem repeats with duplications and excisions of variable degree. IEEE/ACM Trans Comput Biol Bioinform. 2006;3(4):395–407.
37. Mellmann A, Weniger T, Berssenbrügge C, Rothgänger J, Sammeth M, Stoye J, et al. Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of Staphylococcus aureus populations based on spa polymorphisms. BMC Microbiol. 2007; 7:98.
38. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. PHYLOViZ: Phylogenetic Inference and Data Visualization for Sequence Based Typing Methods. BMC Bioinformatics. 2012;13:87.
39. Grundmann H, Klugman KP, Walsh T, Ramon-Pardo P, Sigauque B, Khan W, et al. A framework for global surveillance of antibiotic resistance. Drug Resis Updat. 2011;14(2):79–87.
40. Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, Friedrich AW, et al. Geographic distribution of Staphylococcus aureus causing invasive infections in Europe: a molecular-epidemiological analysis. PLoS Med. 2010;7(1):e1000215.
41. Aanensen DM, Huntley DM, Feil EJ, al-Own F, Spratt BG. EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. PLoS One. 2009;4(9):e6968.
42. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2009;155(2):945–59.
43. Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. Genetics. 2006;175(3):1251–66.

44. Corander J, Marttinen P. Bayesian identification of admixture events using multilocus molecular markers. Mol Ecol. 2006;15(10):2833–43.
45. Corander J, Tang J. Bayesian analysis of population structure based on linked molecular information. Math Biosci. 2007;205(1):19–31.
46. Tang J, Hanage WP, Fraser C, Corander J, Bourne PE. Identifying currents in the gene pool for bacterial populations using an integrative approach. PLoS Comput Biol. 2009; 5(8):e1000455.
47. Dale J, Price EP, Hornstra H, Busch JD, Mayo M, Godoy D, et al. Epidemiological tracking and population assignment of the non-clonal bacterium, Burkholderia pseudomallei. PLoS Negl Trop Dis. 2011;5(12):e1381.
48. Hanage WP, Fraser C, Tang J, Connor TR, Corander J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science. 2009;324(5933):1454–7.
49. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog. 2008;4(9):e1000160.
50. Simpson E. Measurement of diversity. Nature. 1949;163:688.
51. Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol. 2001;39(11):4190–2.
52. Rand WM. Objective criteria for the evaluation of clustering methods. Journal Am Statist Assoc.1971;66:846–50.
53. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193–218.
54. Wallace DL. A method for comparing two hierarchical clusterings: comment. J Am Statist Ass. 1983;78:569–76.
55. Carrico J, Pinto F, Simas C, Nunes S, Sousa N, Frazao N, et al. Assessment of band-based similarity coefficients for automatic type and subtype classification of microbial isolates analyzed by pulsed-field gel electrophoresis. J Clin Microbiol. 2005;43(11):5483–90.
56. Pinto FR, Melo-Cristino J, Ramirez M. A confidence interval for the wallace coefficient of concordance and its application to microbial typing methods. PLoS One. 2008;3(11):e3696.
57. Severiano A, Carriço JA, Robinson DA, Ramirez M, Pinto FR. Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PLoS One. 2011;6(5):e19539.
58. Severiano A, Pinto FR, Ramirez M, Carriço JA. Adjusted Wallace coefficient as a measure of congruence between typing methods. J Clin Microbiol. 2011;49(11):3997–4000.
59. 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.
60. Faria NA, Carriço JA, Oliveira DC, Ramirez M, de Lencastre H. Analysis of typing methods for epidemiological surveillance of both methicillin-resistant and methicillin-susceptible Staphylococcus aureus strains. J Clin Microbiol. 2008;46(1):136–144.
61. Miragaia M, Carrico JA, Thomas JC, Couto I, Enright MC, de Lencastre H. Comparison of Molecular Typing Methods for Characterization of *Staphylococcus epidermidis*: Proposal for Clone Definition. J Clin Microbiol. 2008;46(1):118–29.
62. Berners-Lee T, Hendler J. Publishing on the semantic web. Nature. 2001;410(6832):1023–4.
63. Bizer C, Heath T, Berners-Lee T. Linked data-the story so far. Int J Semantic Web and Inf Systems. 2009; 5(3).
64. Almeida JS, Chen C, Gorlitsky R, Stanislaus R, Aires-de-Sousa M, Eleutério P, et al. Data integration gets 'Sloppy'. Nat Biotechnol. 2006;24(9):1070–1.
65. Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. Nat Rev Genet. 2004;5(3):213–22.
66. Almeida J, Tiple J, Ramirez M, Melo-Cristino J, Vaz C, P Francisco A, Carriço JA. An Ontology and a REST API for equence Based Microbial Typing Data, pp. 21–28. In: Freitas, A, Navarro, A, editors. Bioinformatics for Personalized Medicine. Berlin/Heidelberg:Springer,2012. .
67. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol. 2012;10(9):599–606.
68. Dunne WM, Westblade LF, Ford B. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. Eur J Clin Microbiol Infect Dis. 2012;31(8):1719–26.
69. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. PLoS Pathog. 2012;8(8):e1002824.
70. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N Engl J Med. 2011;365(8):718–24.
71. 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.
72. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med. 2011;365(8):709–17.
73. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open. 2012;2(3).
74. Köser 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.
75. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2012;327(5964):469–74.
76. Croucher N, Harris S, Fraser C, Quail M. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331(6016):430–4.
77. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the southern United States. N Engl J Med.2011;364(17):1626–33.
78. National Food Institute, DTU. Perspectives of a global, real-time microbiological genomic identification system. Brussels: DTU;2011.
79. Palm D, Johansson K, Ozin A, Friedrich AW, Grundmann H, Larsson JT, Struelens MJ. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice, Stockholm, November 2011. Euro Surveill. 2012;17(2):pii=20054. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20054
80. Baker S, Hanage WP, Holt KE. Navigating the future of bacterial molecular epidemiology. Curr Opin Microbiol. 2010;13(5):640-5.

Automated extraction of typing information for bacterial pathogens from whole genome sequence data: *Neisseria meningitidis* as an exemplar

K A Jolley[1], M C Maiden (martin.maiden@zoo.ox.ac.uk)[1]
1. Department of Zoology, University of Oxford, Oxford, United Kingdom

Citation style for this article:
Jolley KA, Maiden MC. Automated extraction of typing information for bacterial pathogens from whole genome sequence data: Neisseria meningitidis as an exemplar. Euro Surveill. 2013;18(4):pii=20379. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20379

Article submitted on 07 December 2012 / published on 24 January 2013

Whole genome sequence (WGS) data are increasingly used to characterise bacterial pathogens. These data provide detailed information on the genotypes and likely phenotypes of aetiological agents, enabling the relationships of samples from potential disease outbreaks to be established precisely. However, the generation of increasing quantities of sequence data does not, in itself, resolve the problems that many microbiological typing methods have addressed over the last 100 years or so; indeed, providing large volumes of unstructured data can confuse rather than resolve these issues. Here we review the nascent field of storage of WGS data for clinical application and show how curated sequence-based typing schemes on websites have generated an infrastructure that can exploit WGS for bacterial typing efficiently. We review the tools that have been implemented within the PubMLST website to extract clinically useful, strain-characterisation information that can be provided to physicians and public health professionals in a timely, concise and understandable way. These data can be used to inform medical decisions such as how to treat a patient, whether to instigate public health action, and what action might be appropriate. The information is compatible both with previous sequence-based typing data and also with data obtained in the absence of WGS, providing a flexible infrastructure for WGS-based clinical microbiology.

Introduction

The application of whole genome sequencing (WGS) technology to clinical microbiology has been described as revolutionary: the opportunities are certainly immense, but so too are the challenges of implementing this technology effectively [1]. Above all, clinical microbiology and epidemiology are pragmatic sciences, which require accurate and understandable information to be delivered to those who need to make medical judgements in real time. Often these judgements have to be made in the absence of complete information, and it is essential that widely understood, accepted and reproducible typing methods are employed to guide these decisions [2]. Just as the advent of molecular techniques challenged phenotypic methodologies over a decade ago – replacing imperfect but at least widely accepted techniques with a plethora of non-standardised alternatives [3] – the high volumes of sequence data have to be carefully managed if they are to provide enlightenment rather than confusion.

The multilocus sequence typing (MLST) paradigm was established in 1998 [4], a time when molecular techniques were beginning to be widely used in the clinical laboratory, but when there was no universally agreed way forward [5]. It was intended as a standardised, reproducible and portable approach that could replace and enhance previous methods, particularly multilocus enzyme electrophoresis (MLEE) [6]. MLST was the first sequence-based approach to the genome-wide characterisation of bacterial isolates to be widely adopted and automated methods for performing the reactions and extracting the sequence information have subsequently been developed [7-9]. At the time MLST was introduced, it was impractical to sequence whole genomes on very large numbers of isolates and early analyses showed that in many cases this was not required. The first MLST scheme, for example, was designed to identify major clones within populations of *Neisseria meningitidis*, the meningococcus, and was able to do this reliably and reproducibly with just seven gene fragments, totalling only 3,284 bp or about 0.15% of the whole genome [10,11]. Similar numbers and sizes of loci have been successful for MLST schemes covering a wide range of organisms, which is an indication of the high degree of structuring present in many bacterial populations. For many bacteria, including the meningococcus, the extent of genetic diversity present even in this small number of genes under stabilising selection is extensive [12]: as of November 2012, each of the gene fragments used as meningococcal MLST loci had between 424 to 675 distinct alleles recorded on the PubMLST *Neisseria* website [13], with 54–94% (mean: 71%) sites that were polymorphic. Furthermore, in the representative *abcZ* locus, all four bases were present at a given site over the known population in 54/433 (12%) of the nucleotide positions (Figure 1). Much of this variation is at low frequency and transitory, but

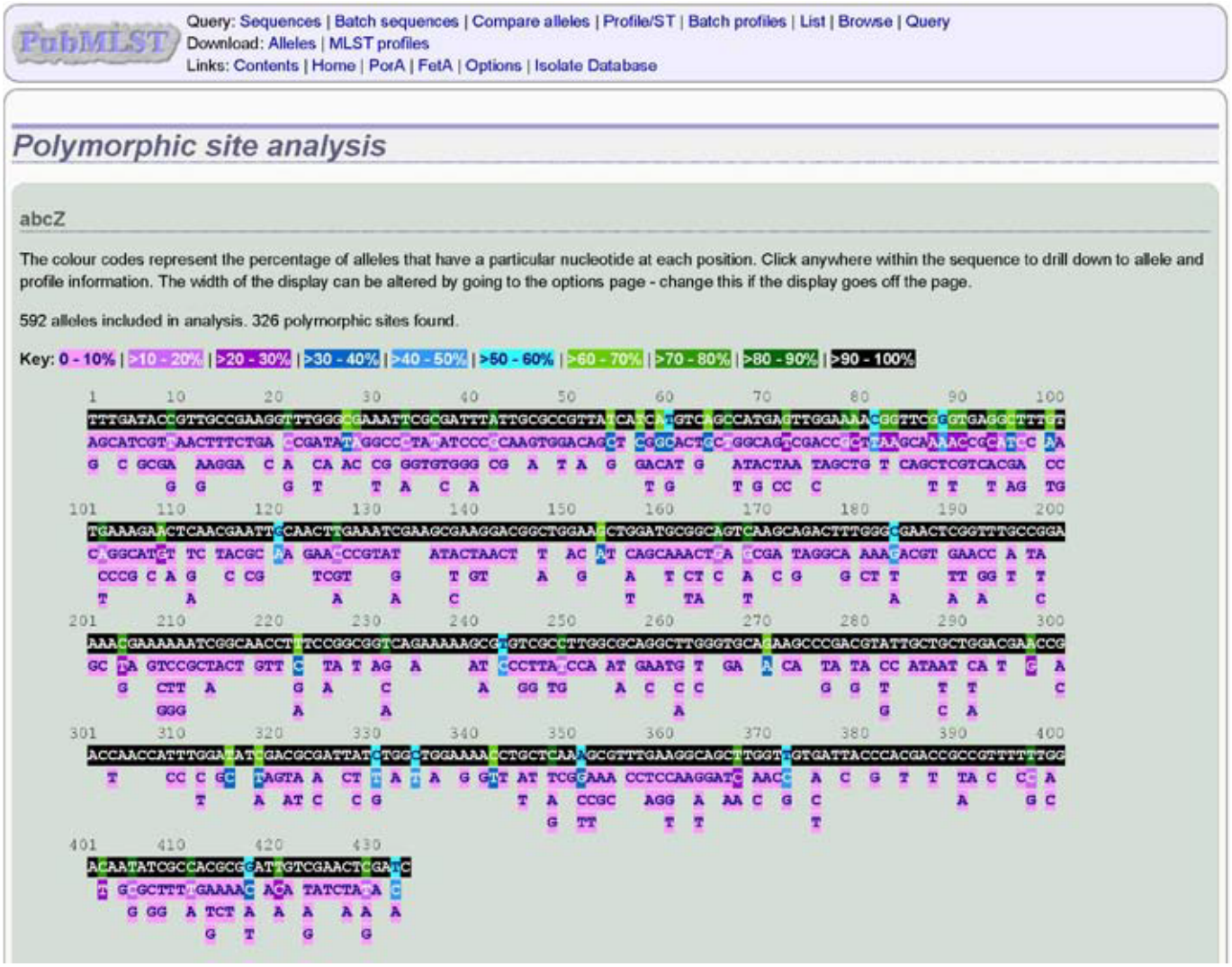
the variants for which this is the case for cannot be known without exhaustive, or at least extensive, sampling over time.

The MLST approach catalogues this extreme diversity, which is seen in many microbial populations and which remains only partially explored, by the maintenance of curated libraries of allele sequences for each MLST locus. Each unique sequence (allele) is assigned a unique arbitrary number, effectively compressing 400–600 bp of information into a single integer. Further organisation and compression of genetic variation is attained by combining the data from all MLST loci into allelic profiles or sequence types (STs), which are also assigned arbitrary numeric designations, each of which defines a unique string of several thousand nucleotides [12]. This approach has proved to be both efficient and effective: as of November 2012, there were 9,927 STs in the *Neisseria* MLST database, for example, each precisely characterising a particular seven-locus *Neisseria* genotype. Similar levels of diversity have been observed in other bacteria hosted at PubMLST and on other MLST repositories [14]. The fact that nearly 10,000 distinct variants of only 3,284 bp of coding sequence under stabilising selection are known to exist in one human-associated bacterium with a genome of about 2.2 Mbp indicates the scale of the cataloguing problem facing us in the era of genomic microbiology.

Nevertheless, there are instances when even the very high levels of diversity routinely seen in MLST datasets do not provide sufficient information for clinical decision-making. This is because even populations of

Figure 1

Schematic of one of the *Neisseria meningitidis* MLST loci (*abcZ*) showing the number and positions of known polymorphic sites within the gene fragment (unmodified PubMLST.org screenshot)



MLST: multilocus sequence typing.
Source: PubMLST *Neisseria* website [13].

diverse organisms, such as the meningococcus, are highly structured, with most isolates belonging to clonal complexes of related bacteria, many of which share identical STs [15]. This detection of population structuring is one of the strengths of the MLST approach, as these clusters are frequently associated with phenotypes of clinical interest such as virulence or expression of vaccine antigens [16]. This clustering, however, can mean that isolates with the same ST may not have the same point source, so ST alone is insufficient to unambiguously identify strains belonging to an outbreak. For this reason, additional highly variable antigenic loci are included in the recommended typing scheme for meningococci [17] and for other organisms such as Campylobacter [18] that are regularly typed by MLST. For meningococci, there are also curated sequence-based schemes for genes that encode antimicrobial resistance that provide additional clinically valuable information [19,20]. Other schemes, such as variable-number tandem repeat (VNTR), also allow high discrimination of isolates in outbreak situations [21,22]. Combining these high-resolution typing approaches with seven-locus MLST and spatial and temporal epidemiology techniques permits the proactive identification of outbreaks of infectious disease [23].

For a small number of bacteria, the so-called single clone pathogens, there is insufficient variation in seven-locus MLST to provide epidemiological resolution, usually because these pathogens have evolved recently from single clones, undergo little recombination and contain too little genetic variation [24]. These include organisms of great medical importance such *Mycobacterium tuberculosis* [25], *Yersinia pestis* [26], *Bacillus anthracis* [27] and *Salmonella enterica* var Typhi [28]. For these bacteria, data from the whole genome, often in the form of single nucleotide polymorphisms (SNPs) [29], but also including other types of variation such as VNTRs, is essential for epidemiological purposes. These data will also have to be stored and interpreted in an accessible way that produces data usable by clinical decision-makers and which is both forwards and backwards compatible.

One of the motivations that drove the development of MLST was future-proofing. Even at a time when the costs of sequencing were seen by some as prohibitive [30], nucleotide sequence data had major advantages: they might be added to, but they would never become obsolete – as they represented the fundamental level of genetic information – and they are readily understood, stored, compared and distributed [12]. Obtaining WGS data is now becoming so inexpensive that it is becoming the fastest and most economical way of obtaining information at multiple loci for determining MLST or other STs [31]. When used in this way, these data are directly comparable to the extensive sequence databases that have been established since the first use of MLST [32,33]. Here we describe how the suite of databases hosted at PubMLST [34] has been updated to accommodate WGS data and describe the tools that are available to rapidly extract typing information from such data. We also describe how these tools can be exploited further to achieve very high resolution from such data when required.

Database structure

As of November 2012, the majority of the typing databases hosted at PubMLST [34] were using the Bacterial Isolate Genome Sequence Database (BIGSdb) platform to archive isolate and sequence diversity data [35]. This software was developed to facilitate the flexible storage and exploitation of the whole range of sequence data that might be available from a clinical specimen, from single Sanger sequencing reads through to whole genomes, which may be either complete or consisting of multiple contiguous sequences ('contigs'), as assembled from data from the current generation of sequencing instruments. The BIGSdb platform consists of two kinds of database: (i) a definition database that contains the sequences of known alleles of loci under study, as well as allelic profiles (combinations of alleles at specific loci) for schemes such as MLST; and (ii) an isolate database that contains isolate provenance and other metadata along with nucleotide sequences associated with that isolate. An isolate database can interact with any number of definition databases and vice versa, allowing networks of authoritative nomenclature servers and partitioning of isolate datasets and projects, with curator access controlled by specific permissions set by an administrator.

Reference databases

The definition databases are central to genome analysis using the gene-by-gene (MLST-like) analysis approach implemented in BIGSdb. By storing all known allelic diversity for any locus of interest, the definition databases provide a centralised queryable repository that provides a common language for expressing sequence differences, making it a trivial process to identify alleles that are different among isolates, and equally importantly, those that are identical. Because sequence differences are linked directly to a particular locus (which can be any definable sequence string, nucleotide or peptide) and with appropriate grouping of loci into 'schemes' (groups of related loci), the context of this locus is immediately apparent: identifying it, for example, as a member of a conventional MLST scheme, as responsible for antimicrobial resistance, as a participant of a biochemical pathway and so on. As of November 2012, the *Neisseria* PubMLST definition database had allelic sequences defined for 1,272 loci with 114,469 unique alleles.

Extracting typing information

Web-based and stand-alone tools have been developed that facilitate identification of STs directly from short-read data [36,37]. These methods are, of course, dependent on the sequence and profile definitions made available on PubMLST.org, which also has functionality to extract typing information directly from

submitted assembled genomes that are routinely scanned for known alleles. As the locations of these loci are 'tagged' in the sequence data for future reference within BIGSdb, this means that the genome sequences are automatically annotated for those loci for which definition databases exist. The definition database can also be queried using genome data not uploaded to the isolate database to identify a strain directly from sequence data. The BIGSdb platform also has functionality that enables an administrator to define scanning rules and report formatting. This uses a built-in script interpreter so that analysis paths can be taken by following a decision tree defined by the rules. This has been implemented within the PubMLST *Neisseria* sequence definition database to automatically extract the strain typing information for the meningococcus (ST, clonal complex and antigen sequence type comprising PorA variable regions and FetA variable region) [17,33], along with antibiotic resistance information from sequence data that is pasted in to a web form (Figure 2, panel A). The script instructs the software to first scan the MLST alleles and, if these are all identified, to identify the ST and clonal complex by querying the reference data tables. It then scans the typing antigens and formats the results of these with the MLST results in to a standardised strain designation [17]. Following this, the sequences of the penA and rpoB genes are extracted and then compared with isolates with matching sequences within the PubMLST isolate database to determine the most likely penicillin and rifampicin sensitivity. All of this is displayed in a plain language report (Figure 2, panel B). The whole analysis is extremely rapid, taking about 40 seconds within the web interface.

Comparing genomes

Because genomic diversity is recorded within BIGSdb as allele numbers, WGS analysis is possible using the highly scalable techniques developed for seven-locus MLST. Once loci have been defined and alleles identified, they can be used essentially as a whole-genome MLST scheme, or any chosen subset of predefined loci combined to form a scheme. This is the principle behind the Genome Comparator analysis [38], which can use either the defined loci or extract coding sequences from an annotated reference genome to perform comparisons against genomes within the database. Using a reference genome, or set of predefined reference loci, each of the coding sequences are compared against the test genomes using BLAST. Allele sequences that are the same as the reference are designated allele 1, while each unique allele different from the reference is assigned a sequential number. Once each locus has been tested, a distance matrix is then generated based on allelic identities between each pair of isolates. This can then be visualised using standard algorithms – the PubMLST website incorporates the Neighbor-net algorithm [39] implemented in SplitsTree4 [40]. Because analysis relies only on using BLAST to compare each locus within a genome in turn, either against the single annotated reference sequence or against all known alleles if using defined loci, the analysis is again very rapid, allowing multiple genomes to be compared within minutes, with the time taken to analyse only increasing linearly, not geometrically, with additional genomes.

The Genome Comparator approach is generic and any number of loci in any groups can be used for this type of analysis. Many loci have been defined for the meningococcus, including the 53 ribosomal (r) genes that are used as a basis of rMLST [41-44]. The full complement of ribosomal genes has a number of advantages for indexing variation. These genes are universally present in members of the domain, are protein encoding and therefore generally assemble well from short-read sequences and are distributed throughout the genome. They encode proteins that form part of a coherent, macromolecular structure and contain variation that is informative at a wide range of levels of discrimination. These data can be used within and among members of the same genus, for both species and strain definition [42].

Analysis of whole genome sequence data for meningococci

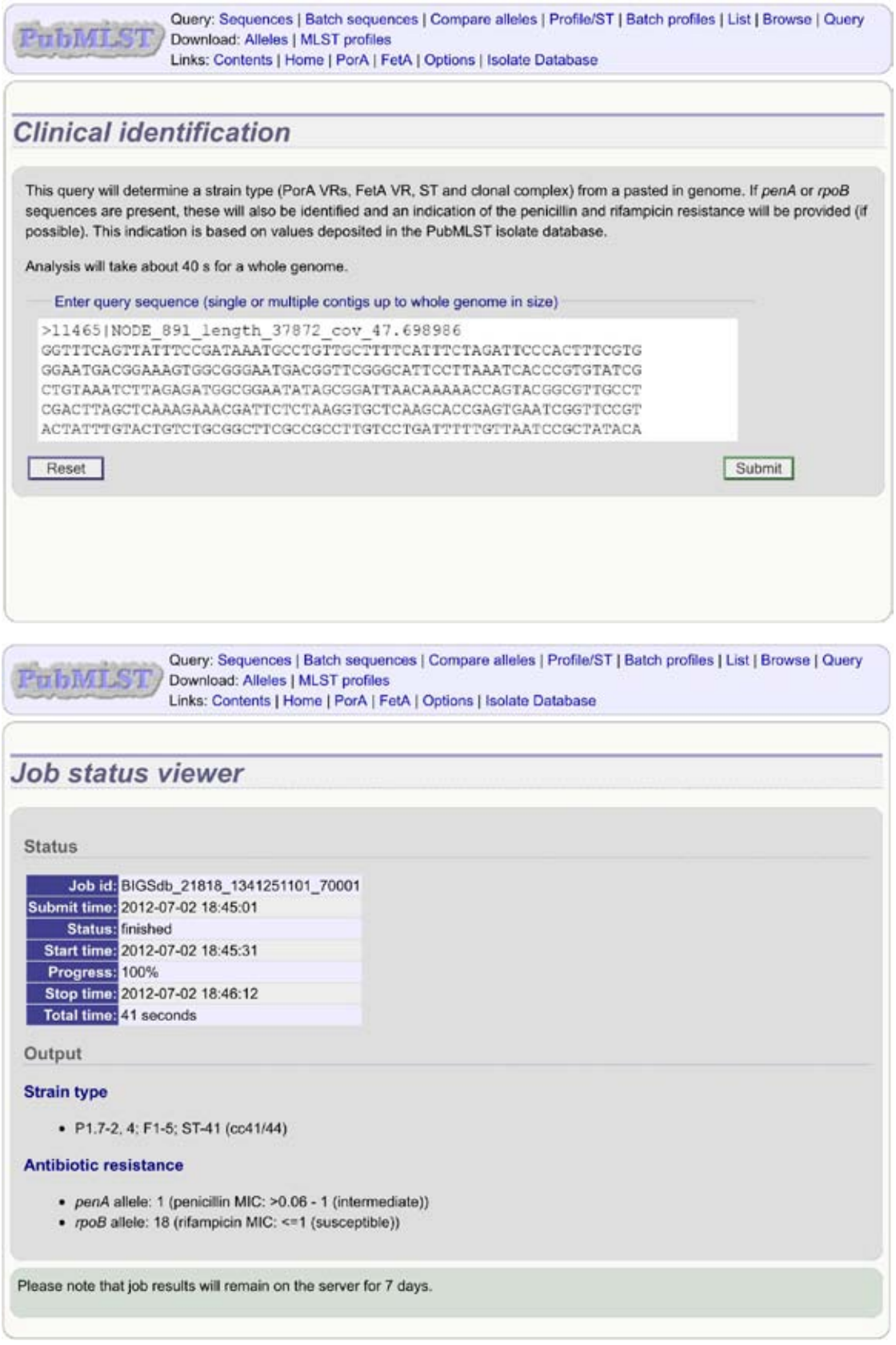
The *Neisseria* PubMLST database is continually expanding: as of November 2012, there were 221 isolate records with deposited genome sequence data linked to published studies [11,45-51]. Of these 221 genomes, 170 were meningococci, with the remainder belonging to other species within the genus [42]. The data consisted of a mixture of finished genomes, multiple contigs generated from de novo assembly, contigs generated by mapping to a reference sequence and sets of predicted coding sequences. These are treated identically by BIGSdb to identify and tag sequences of known loci, and where these loci are members of existing typing schemes, such as MLST or antigen typing, these genomes could be compared to legacy data (Table).

Neighbor-net visualisation of distance matrices generated with Genome Comparator from allelic rMLST data [44] provides a highly scalable, rapid and easily understood way of placing isolates within the known diversity of a bacterial species. For example, the interrelationships of 139 *N. meningitidis* isolates present in the PubMLST *Neisseria* database [13] can be efficiently represented by this method. Since rMLST alleles are automatically tagged within the database, this analysis is rapid and the Neighbor-net trees can be generated in a few minutes. The rMLST analysis differentiates clonal complexes; however, in addition it provides much higher resolution than conventional seven-locus MLST [38], robustly indicating both relationships among and diversity within clonal complexes (Figure 3).

The locations of isolates belonging to major clonal complexes identified by conventional MLST are indicated (cc1, etc.). The figure illustrates relationships not apparent from seven-locus MLST, including the

Figure 2

Extracting antigen and antibiotic resistance data from *Neisseria meningitidis* whole genome sequences



A whole genome sequence, which may consist of multiple contigs, can be pasted in to the Neisseria PubMLST website (panel A) with typing and antibiotic resistance data for penicillin and rifampicin rapidly extracted (panel B) (unmodified PubMLST.org screenshots).

Source: PubMLST Neisseria website [13].

Table

Meningococcal whole genome sequencing data linked to published studies, deposited in the PubMLST *Neisseria* database as of November 2012

Clonal complex	Number of genome sequences	Number of STs	Serogroups	PorA variant combinations	FetA variants
cc11	31	6	C (22), W (4), B(2), NG (1), NA (2)	8	8
cc41/44	20	12	B (14), NA (5), NG (1)	10	5
cc32	17	4	B (14), C (1), NG (1), NA (1)	10	5
cc5	16	5	A (16)	3	5
cc4	14	1	A (14)	4	1
cc1	13	3	A (13)	4	5
cc8	9	5	B (5), C (3), NA (1)	6	5
cc18	5	4	B (4), C (1)	5	4
cc23	5	2	Y (5)	3	2
cc22	4	1	W (4)	1	2
cc167	4	4	Y (4)	1	2
cc269	4	3	B (2), NA (2)	4	3
cc37	2	2	B (2)	1	2

NA: not available; NG: non-groupable; ST: sequence type.

The table shows the clonal complex and indicates the diversity of ST, serogroup and typing antigens. Only clonal complexes represented by two or more genomes are included.

diversity of some clonal complexes (e.g. cc1) and the interrelationships of others, e.g. cc8 and cc11 clonal complexes, and the relationships of the ET-15 and ET-37 variants within cc11.

Conclusions and future prospects

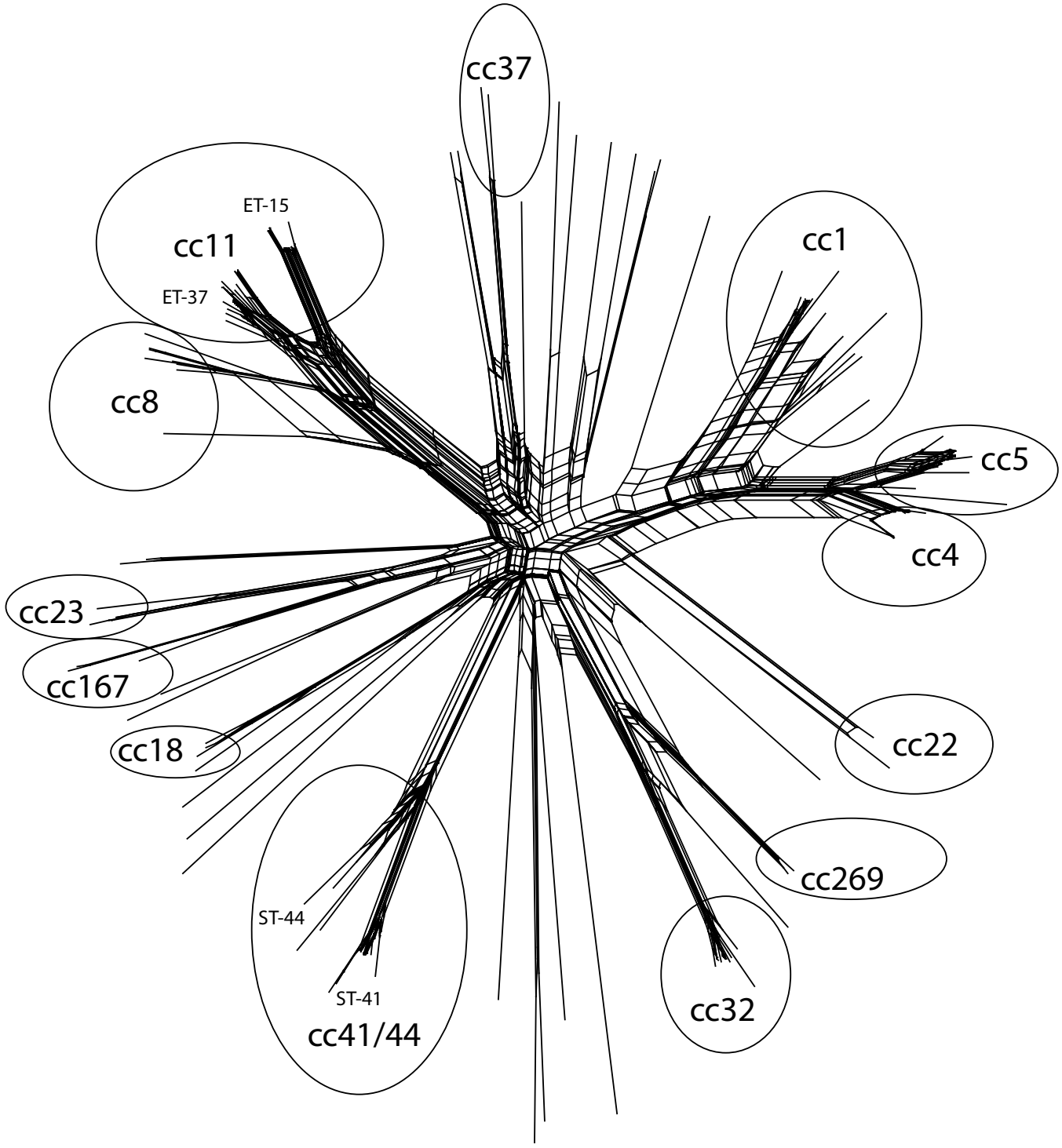
Nucleotide sequences are a universal language that can be interpreted in a number of ways. For clinical and epidemiological purposes, sequences from clinical specimens have to be rapidly and effectively translated into a meaningful term or set of terms that define those properties of the aetiological agents of disease that direct medical and public health action. One of the factors behind the success of seven-locus MLST was the introduction of standard sets of nomenclature that reflected the structure of microbial populations and their phenotypic properties. For organisms with well-established and accepted MLST and other typing schemes in place, the impact of the application of WGS data will be to rapidly identify properties such as strain type. In some cases, novel nomenclature may be required, but this is a process that has to be approached with care, if confusion in the wider clinical community is to be avoided.

The suite of database subsites on PubMLST, which now includes a site that catalogues the ribosomal diversity across the whole domain for the purposes of rMLST typing [44,52], provides an example of how WGS data can be used to efficiently designate specimens to current strain types. It can be also used to establish additional typing schemes which can coexist with each other side by side, as there is no limit to the number of loci and schemes that can be defined. As the database stores the sequence information that is available for an isolate, be that a single read or a whole genome, it means that it is possible to seamlessly compare isolates for which different types of information are available, achieving backwards compatibility with previous typing schemes, as well as compatibility with diagnostic tests that may target only one or a few loci. The extent to which isolates can be compared depends only on the quality of the sequence data available for the locus in question, but given that clinical specimens are often imperfect, it is important for clinical and epidemiological purposes that incomplete or partial information can be used. While many studies place short-read data in a sequence read archive, this is not easily accessible or readily analysed. PubMLST curators do proactively assemble short-read data and incorporate the resultant contigs into the database where metadata are available. Links are made to the sequence read archive within PubMLST isolate records so that original data can be retrieved and analysed when required. While the *Neisseria* databases described are exemplars, databases for other species can be hosted on request and the open-source BIGSdb software is freely available for local installation.

The first analyses of WGS data on bacterial specimens relied on SNP analysis of closely related bacteria, with mapping of sequence reads to a predefined reference genome. These have required pre-analysis of the samples by an approach such as MLST to limit the extent

Figure 3

Relationships of 139 *Neisseria meningitidis* genomes in the PubMLST *Neisseria* database, generated with Genome Comparator and Neighbor-net from allelic profiles data for rMLST loci



r: ribosomal; MLST: multilocus sequence typing.

The locations of isolates belonging to major clonal complexes identified by conventional MLST are indicated (cc1, etc.). The figure illustrates relationships not apparent from seven-locus MLST, including the diversity of some clonal complexes (e.g. cc1) and the interrelationships of others, e.g. cc8 and cc11 clonal complexes, and the relationships of the ET-15 and ET-37 variants within cc11.

of variation being analysed [53-58]. This approach is also appropriate and can be very effective for 'single clone' pathogens [25-28]; however, it is not feasible for the general analysis for diagnosis or surveillance of bacteria such as the meningococcus that exhibit more typical levels of sequence diversity. Indeed, the use of the term SNP when discussing bacterial genome variation outside the examples described above, is unfortunate and can be misleading. The concept of the 'SNP' has been taken from human medical genomics to microbial genomics: in humans, it is in some cases appropriate to discuss SNPs, when they are associated with a particular genetic disease, but genetic variation in terms of sequence polymorphism is much more complex in bacteria. As seen here, the great majority of microbial populations contain tens of thousands of polymorphisms even within organisms that are closely related – not to mention large amounts of variation due to insertions, deletions and rearrangements, which cannot even remotely be described as 'SNPs'. The term sequence variation is more appropriate as individual polymorphisms, especially in bacteria, are invariably embedded with many other variants into alleles and it is these alleles – each often with many variable sites – that are associated with particular phenotypes.

Although the typing of bacterial specimens with existing schemes is a valuable contribution of WGS data to clinical microbiology and epidemiology, it is not, of course, the only use for these data. There are many other possible applications for both research and detailed investigation of outbreaks [38]; however, it is important that the analysis of these data is driven by the question that is being asked. If an outbreak can be resolved with a few loci, then there is no need to pursue the data further and certainly no need to report more detail than necessary to a hard-pressed front-line clinician or epidemiologist who, in general, will only require the information necessary to resolve the medical problem at hand. In other cases, resolution of a particular outbreak may require data from the whole genome [53]. For this reason, it will be increasingly necessary to store WGS data from clinical specimens in an understandable form, that is, as assembled sequences, within flexible structures, such as that offered by the PubMLST platform powered by BIGSdb, where WGS information can be hierarchically queried in real time by individuals with limited bioinformatics expertise to generate the data at the resolution required to address their problem. In this context these data will provide an exciting opportunity to extend our understanding of infectious disease caused by bacteria and will enhance our ability to combat it.

References

1. Pallen MJ, Loman NJ, Penn CW. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr Opin Microbiol. 2010;13(5):625-31.
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. Achtman M. A surfeit of YATMs? J Clin Microbiol. 1996;34(7):1870.
4. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA. 1998;95(6):3140-5.
5. van Belkum A, Struelens M, de Visser A, Verbrugh H, Tibayrenc M. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin Microbiol Rev. 2001;14(3):547-60.
6. 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):837-84.
7. Sullivan CB, Jefferies JM, Diggle MA, Clarke SC. Automation of MLST using third-generation liquid-handling technology. Mol Biotechnol. 2006;32(3):219-26.
8. Platt S, Pichon B, George R, Green J. A bioinformatics pipeline for high-throughput microbial multilocus sequence typing (MLST) analyses. Clin Microbiol Infect. 2006;12(11):1144-6.
9. O'Farrell B, Haase JK, Velayudhan V, Murphy RA, Achtman M. Transforming microbial genotyping: a robotic pipeline for genotyping bacterial strains. PLoS One. 2012;7(10):e48022.
10. Holmes EC, Urwin R, Maiden MC. The influence of recombination on the population structure and evolution of the human pathogen Neisseria meningitidis. Mol Biol Evol. 1999;16(6):741-9.
11. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, et al. Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491. Nature. 2000;404(6777):502-6.
12. Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol. 2006;60:561-88.
13. Neisseria sequence typing home page. Oxford: University of Oxford. [Accessed 31 Nov 2012]. Available from: http://pubmlst.org/neisseria/
14. All species MLST databases and published schemes. [Accessed 31 Nov 2012]. Available from: http://pubmlst.org/databases.shtml
15. Caugant DA, Maiden MC. Meningococcal carriage and disease--population biology and evolution. Vaccine. 2009;27 Suppl 2:B64-70.
16. Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, et al. Distribution of serogroups and genotypes among disease-associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway. J Clin Microbiol. 2004;42(11):5146-53.
17. Jolley KA, Brehony C, Maiden MC. Molecular typing of meningococci: recommendations for target choice and nomenclature. FEMS Microbiol Rev. 2007;31(1):89-96.
18. Dingle KE, McCarthy ND, Cody AJ, Peto TE, Maiden MC. Extended sequence typing of Campylobacter spp., United Kingdom. Emerg Infect Dis. 2008;14(10):1620-2.
19. Taha MK, Hedberg ST, Szatanik M, Hong E, Ruckly C, Abad R, et al. Multicenter study for defining the breakpoint for rifampin resistance in Neisseria meningitidis by rpoB sequencing. Antimicrob Agents Chemother. 2010;54(9):3651-8.
20. Taha MK, Vázquez JA, Hong E, Bennett DE, Bertrand S, Bukovski S, et al. Target gene sequencing to characterize the penicillin G susceptibility of Neisseria meningitidis. Antimicrob Agents Chemother. 2007;51(8):2784-92.
21. Schouls LM, van der Ende A, Damen M, van de Pol I. Multiple-locus variable-number tandem repeat analysis of Neisseria meningitidis yields groupings similar to those obtained by multilocus sequence typing. J Clin Microbiol. 2006;44(4):1509-18.
22. Elias J, Schouls LM, van de Pol I, Keijzers WC, Martin DR, Glennie A, et al. Vaccine preventability of meningococcal clone, Greater Aachen Region, Germany. Emerg Infect Dis. 2010;16(3):464-472.
23. Elias J, Harmsen D, Claus H, Hellenbrand W, Frosch M, Vogel U. Spatiotemporal analysis of invasive meningococcal disease, Germany. Emerg Infect Dis. 2006;12(11):1689-95.
24. Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol. 2008;62:53-70.

25. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathogens. 2008;4(9):e1000160.
26. Haensch S, Bianucci R, Signoli M, Rajerison M, Schultz M, Kacki S, et al. Distinct clones of *Yersinia pestis* caused the Black Death. Plos Pathogens. 2010;6(10):e1001134.
27. Pearson T, Okinaka RT, Foster JT, Keim P. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect Genet Evol. 2009;9(5):1010-9.
28. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet. 2008;40(8):987-93.
29. Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis. 2004;10(9):1568-77.
30. Olive DM, Bean P. Principles and applications of methods for DNA-based typing of microbial organisms. J Clin Microbiol. 1999;37(6):1661-9.
31. Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, et al. Microbiology in the post-genomic era. Nat Rev Microbiol. 2008;6(6):419-30.
32. 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.
33. Vogel U, Szczepanowski R, Claus H, Jünemann S, Prior K, Harmsen D. Ion torrent personal genome machine sequencing for genomic typing of Neisseria meningitidis for rapid determination of multiple layers of typing information. J Clin Microbiol. 2012;50(6):1889-94.
34. PubMLST. Oxford: University of Oxford. [Accessed 31 Nov 2012]. Available from: http://pubmlst.org/
35. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.
36. 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.
37. Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics. 2012;13:338.
38. Jolley KA, Hill DM, Bratcher HB, Harrison OB, Feavers IM, Parkhill J, et al. Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid web-based analysis methods. J Clin Microbiol. 2012;50(9):3046-53.
39. 39. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004;21(2):255-65.
40. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67.
41. Read DS, Woodcock DJ, Strachan NJ, Forbes KJ, Colles FM, Maiden MC, et al. Evidence for phenotypic plasticity among multihost Campylobacter jejuni and C. coli lineages, obtained using ribosomal multilocus sequence typing and Raman spectroscopy. Appl Environ Microbiol. 2013;79(3):965-73.
42. Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD, Parkhill J, Maiden MC. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria. Microbiology. 2012;158(Pt 6):1570-80.
43. Ussery DW, Gordon SV. Two novel methods for using genome sequences to infer taxonomy. Microbiology. 2012;158(Pt 6):1414.
44. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony CM, Colles FM, et al Ribosomal multilocus sequence typing: universal characterisation of bacteria from domain to strain. Microbiology. 2012;158(Pt 4):1005-15.
45. Hao W, Ma JH, Warren K, Tsang RS, Low DE, Jamieson FB, et al. Extensive genomic variation within clonal complexes of Neisseria meningitidis. Genome Biol Evol. 2011;3:1406-18.
46. Budroni S, Siena E, Hotopp JCD, Seib KL, Serruto D, Nofroni C, et al. Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc Natl Acad Sci USA. 2011;108(11):4494-9.
47. Schoen C, Blom J, Claus H, Schramm-Glück A, Brandt P, Müller T, et al. Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in Neisseria meningitidis. Proc Natl Acad Sci USA. 2008;105(9):3473-8.
48. Katz LS, Humphrey JC, Conley AB, Nelakuditi V, Kislyuk AO, Agrawal S, et al. Neisseria Base: a comparative genomics database for Neisseria meningitidis. Database (Oxford). 2011;2011:bar035.
49. Rusniok C, Vallenet D, Floquet S, Ewles H, Mouzé-Soulama C, Brown D, et al. NeMeSys: a biological resource for narrowing the gap between sequence and function in the human pathogen Neisseria meningitidis. Genome Biol. 2009;10(10):R110.
50. Bentley SD, Vernikos GS, Snyder LA, Churcher C, Arrowsmith C, Chillingworth T, et al. Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. PLoS Genet. 2007;3(2):e23.
51. Peng J, Yang L, Yang F, Yang J, Yan Y, Nie H, et al. Characterization of ST-4821 complex, a unique Neisseria meningitidis clone. Genomics. 2008;91(1):78-87.
52. Ribosomal multilocus sequence typing (rMLST). Oxford: University of Oxford. [Accessed 31 Nov 2012]. Available from: http://rmlst.org/
53. Köser 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.
54. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331(6016):430-4.
55. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010;327(5964):469-74.
56. 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 USA. 2012;109(8):3065-70.
57. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, et al. Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. Proc Natl Acad Sci USA. 2012;109(12):4550-5.
58. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open. 2012;2(3). pii: e001124.

Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012

B A Lindstedt (bjorn-arne.lindstedt@fhi.no)[1], M Torpdahl[2], G Vergnaud[3,4], S Le Hello[5], F X Weill[5], E Tietze[6], B Malorny[7], D M Prendergast[8], E Ní Ghallchóir[8], R F Lista[9], L M Schouls[10], R Söderlund[11], S Börjesson[11], S Åkerström[11]

1. Division of Infectious Diseases Control, Norwegian Institute of Public Health, Oslo, Norway
2. Department of Microbiological Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark
3. Université Paris-Sud, Institut de Génétique et Microbiologie, Unités Mixtes de Recherche (UMR) 8621, Orsay, France
4. Direction Générale de l'Armement (DGA)/Mission pour la Recherche et l'Innovation Scientifique (MRIS), Bagneux, France
5. Institut Pasteur, Unité de Recherche et d'Expertise des Bactéries Pathogènes Entériques, Centre National de Référence E. coli/Shigella/Salmonella, Paris, France
6. National Reference Center for Salmonella and other Enterics, Robert Koch Institute, Wernigerode Branch, Wernigerode, Germany
7. Federal Institute for Risk Assessment (BfR) National Salmonella Reference Laboratory Department Biological Safety, Berlin, Germany
8. Central Veterinary Research Laboratory, Department of Agriculture, Food and the Marine, Kildare, Ireland
9. Health Corps Italian Army, Department of Molecular Biology, Immunology and Experimental Medicine, Army Medical and Veterinary Research Center, Rome, Italy
10. Laboratory for Infectious Diseases and Perinatal Screening (LIS), Centre for Infectious Disease Control Netherlands (CIb), National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
11. The National Veterinary Institute (SVA), Uppsala, Sweden

Citation style for this article:
Lindstedt BA, Torpdahl M, Vergnaud G, Le Hello S, Weill FX, Tietze E, Malorny B, Prendergast DM, Ní Ghallchóir E, Lista RF, Schouls LM, Söderlund R, Börjesson S, Åkerström S. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill. 2013;18(4):pii=20385. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20385

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

Genotyping of important medical or veterinary prokaryotes has become a very important tool during the last decades. Rapid development of fragment-separation and sequencing technologies has made many new genotyping strategies possible. Among these new methods is multilocus variable-number tandem repeat analysis (MLVA). Here we present an update on the use of MLVA in eight European countries (Denmark, France, Germany, Ireland, Italy, the Netherlands, Norway and Sweden). Researchers in Europe have been active in developing and implementing a large array of different assays. MLVA has been used as a typing tool in several contexts, from aiding in resolving outbreaks of foodborne bacteria to typing organisms that may pose a bioterrorist threat, as well as in scientific studies.

Introduction

Multilocus variable-number tandem repeat analysis (MLVA) is a DNA-based molecular typing method frequently applied to the study of prokaryotes. It records size polymorphisms in several variable-number of tandem repeats (VNTR) loci amplified by stringent PCR protocols. MLVA will mainly impact the public health field by introducing newer, faster and safer (reduced handling of live bacteria) methodologies for typing microorganisms. Reduced typing time, with high resolution, is beneficial for resolving large and complex outbreak situations. The methodology is also suitable for large-scale automation: suitable instruments (e.g. automated sequencers, pipetting robots and analytical software) are already commercially available. There are several variations of MLVA assays depending on available instrumentation. Earlier versions tended to measure VNTR sizes by agarose gel electrophoresis, while newer assays often use capillary electrophoresis for size determination once the allele size range at each locus has been well characterised.

As mentioned above, MLVA assays have clear advantages, offering fast typing, high resolution and reduced handling times of pathogenic organisms. Their drawbacks include high assay-specificity (e.g. each organism usually needs a distinct MLVA assay) and the, as yet, lack of standardisation for the majority of published assays. In Europe, only the *Salmonella enterica* subspecies *enterica* serovar Typhimurium (*S.* Typhimurium) MLVA assay has achieved generally accepted standardisation [1,2]. MLVA is gaining in popularity: in 2000, there was only one PubMed entry (when searching for 'MLVA') while in 2011, there were 96 entries for articles that year alone. There has been extensive research on MLVA and MLVA protocol development within Europe: an overview of organisms for which there are existing MLVA assays in European countries, based on web searches for protocols is presented in Table 1. The web searches were performed on 23 April 2012 and repeated on 18 June in PubMed using the search terms; 'MLVA', 'VNTR', 'tandem repeats', 'TR', 'direct repeats', 'DR' and 'genotyping', combined with geographical names such as 'Europe', 'European' or the countries within Europe. General Internet searches using the same keywords in a standard web browser were also included. The same

Table 1

Multilocus variable-number tandem repeat analysis (MLVA) assays used in 17 countries in Europe, 2012

Organism	AT	BE	DK	FR	DE	EL	IE	iT	NL	NO	PL	PT	RU	ES	SE	CH	UK
Acinetobacter baumannii				X				X									
Bacillus anthracis		X		X	X			X			X		X		X		
Bartonella henselae				X							X						
Bordetella pertussis			X						X						X		X
Brucella spp.			X	X	X	X		X	X				X	X		X	
Clostridium botulinum				X				X									
Coxiella burnetii				X	X				X		X				X		X
Clostridium difficile				X	X				X								X
Chlamydia trachomatis				X					X								X
Escherichia coli					X		X			X					X		X
Enterococcus faecium					X				X					X	X	X	
Francisella tularensis				X											X		
Listeria monocytogenes			X				X			X						X	
Legionella pneumophila				X	X			X									
Mycobacterium bovis				X			X	X				X					X
Mycobacterium leprae				X													X
Mycobacterium avium subsp. *paratuberculosis*				X	X				X								
Mycoplasma pneumoniae				X	X									X			
Mycobacterium tuberculosis	X		X	X	X	X		X	X	X	X	X	X	X	X	X	X
Neisseria gonorrhoeae									X				X				
Neisseria meningitidis						X		X	X	X							
Pseudomonas aeruginosa				X					X								
Streptococcus agalactiae				X						X							
Staphylococcus aureus		X		X	X		X	X	X			X		X		X	X
Salmonella enteritidis	X		X		X												X
Staphylococcus epidermidis															X	X	
Streptococcus pneumoniae				X					X		X			X			
Salmonella typhimurium			X	X	X		X		X	X				X	X		X
Shigella spp.				X						X							
Vibrio cholerae										X			X				
Yersinia enterocolitica										X	X						
Yersinia pestis				X	X			X					X				

AT: Austria; BE: Belgium; CH: Switzerland; DE: Germany; DK: Denmark; EL: Greece; ES: Spain; FR: France; IE: Ireland; IT: Italy; NL: Netherlands; NO: Norway; PL: Poland; PT: Portugal; RU: Russia; SE: Sweden; UK: United Kingdom.

X denotes the use of an MLVA assay in the country. The table includes only assays used by more than one country, and countries with published results using more than one MLVA assay. Web searches were performed on 23 April 2012 and 18 June in PubMed and Internet searches (using the search terms 'MLVA', 'VNTR', 'tandem repeats', 'TR', 'direct repeats', 'DR' and 'genotyping', combined with geographical names such as 'Europe', 'European' or the countries within Europe).

searches were also repeated using Google Scholar and the Scirus search engine.

In this Euroroundup, we present a more in-depth update on the use of MLVA in eight European countries. European researchers with publications describing the development or use of MLVA assays were contacted: those who chose to contribute to this Euroroundup were included. The authors were given a choice of writing a general overview of MLVA assays used in their respective countries and/or giving examples where MLVA has been used to improve public health, e.g. by aiding in solving outbreaks.

Denmark

In Denmark, culture-confirmed cases of *Salmonella* and *Listeria* infection are notifiable by clinical laboratories to the Statens Serum Institut (SSI). Furthermore, all isolates are routinely sent to SSI from the local clinical departments and are included in the national surveillance data. All *Listeria* isolates and the two main serotypes of *Salmonella* – *S.* Typhimurium (including the monophasic variant) 4,[5],12:i:- and *Salmonella enterica* subspecies *enterica* serovar Enteritidis (S. Enteritidis) – are real-time typed using MLVA in order to investigate clusters and detect outbreaks.

All incoming *S.* Typhimurium isolates have been typed by MLVA [1] at SSI since 2003 and all MLVA fragments are converted to true allele numbers using the reference collection and standardised MLVA method [2]. As of April 2012, a total of 6,118 *S.* Typhimurium isolates had been MLVA typed for routine surveillance and separated into 1,102 different MLVA types. Several clusters have been investigated in this period [3] and the implementation of MLVA has helped to define and solve both national and international outbreaks [4]. MLVA has furthermore been used for typing of food, feed and animal isolates, enhancing our ability to identify the source of a food-borne outbreak.

Three MLVA types (2-11-13-9-212, 2-15-7-10-212 and 3-20-7-6-212) accounted for more than 28% of all isolates in Denmark and were seen in an outbreak that lasted over two years (2008–2009) and included more than 1,700 patients [5]. The limited number of genotypes identified was not due to a lack of discrimination using MLVA or indeed pulsed-field gel electrophoresis (PFGE) or phage typing: all three methods were applied during this outbreak, which was unfortunately never solved. Several isolates from the entire period that this outbreak took place have undergone whole genome shotgun sequencing: very few single nucleotide polymorphisms (SNPs) are present in these three MLVA types. These data will be presented in a later manuscript.

Another group, accounting for 13% of all *S.* Typhimurium isolates, is comprised of five closely related MLVA types that have been predominant from 2005 and still are (the five types are the constant loci STTR9 (3), STTR10 (NA) and STTR3 (211) and different combinations of the variable loci STTR5 (11,12,13) and STTR6 (9,10), where paranthesised numbers denote allele sizes and NA (no amplification) indicates negative PCR amplification, as previously described [2].

MLVA typing of S. Enteritidis has been carried out for routine surveillance since 2009 [6] and all MLVA fragments are converted to true allele numbers using the reference collection and five standardised loci [7]. By April 2012, a total of 1,371 S. Enteritidis isolates had been MLVA typed and divided into 131 different MLVA types. The Danish routine surveillance MLVA data have been used in defining clusters and linking patients with an S. Enteritidis infection to a common source or event. A high percentage of S. Enteritidis infections in Denmark are acquired abroad and MLVA typing of S. Enteritidis could be of added value when trying to define and solve international outbreaks in the future. Two groups of MLVA types account for more than half of all S. Enteritidis isolates. One group, seen in 33% of isolates, consists of three MLVA types with four loci in common – SE1 (3), SE2 (7), SE9 (2) and SE3 (4) – and one variable locus, SE5 (10, 12 or 13). Two MLVA types make up 25% and have four loci in common – SE1 (4), SE2 (5), SE9 (3) and SE3 (3) – and one variable locus, SE5 (9 or 10).

For molecular surveillance of *Listeria* infections, SSI uses an in-house developed MLVA method that has shown promise in cluster detection and outbreak investigations. The method is still being validated in our laboratory by comparing MLVA data with those from PFGE.

France

French researchers have been very active for more than 10 years in developing MLVA for the genotyping of pathogenic bacteria and fungi of global health interest (concerning humans, animals and plants) or which may pose a bioterrorist threat. These developments have included the setting up of new assays and of tools accessible on the Internet to facilitate the development of such assays [8]. Of particular interest are online databases presenting MLVA typing data, including the first one, made public in 2002 [8], the development and commercialisation of typing kits and the provision of typing services. MLVA is currently in the phase of entering routine practice in a number of reference laboratories and a market seems to be emerging in France.

MLVA is primarily used in France for six bacterial species of high medical interest. The MLVA used for Mycobacterium tuberculosis [9] is now well-known worldwide as mycobacterial interspersed repetitive units- variable-number tandem repeat (MIRU-VNTR), owing to the efforts of a company (Genoscreen, Lille, France) in Institut Pasteur Lille and to the importance of this pathogen. This assay has also served as a pilot for the development of large-scale MLVA typing and associated databases. More recently, MLVA has been developed for *Staphylococcus aureus*, Legionella

pneumophila and Pseudomonas aeruginosa, with the production of fully automated assays and of typing kits by the Centre Européen d'Expertise et de Recherche sur les Agents Microbiens (CEERAM) at La Chapelle sur Erdre. In the *L. pneumophila* assay, 12 loci are co-amplified in a single multiplex PCR [10]. Alternatively, the assays can be set up locally, with no need to buy kits, since all the necessary information is published [10-12]. MLVA is also in routine use for Streptococcus pneumoniae, with more than 1,000 genotypes publicly accessible from the the Robert Picqué Military Hospital in Bordeaux [13] and for Acinetobacter baumanii [14].

An MLVA assay for *Streptococcus agalactiae* has also been developed in France and additional MLVA assays are currently being developed by the Agence nationale de sécurité sanitaire de l'alimentation (ANSES) for zoonotic agents and by the Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) for plant pathogens.

MLVA assays, which are now used worldwide, have also been developed for major bioterrorist agents, including *Yersinia pestis* and *Bacillus anthracis* [15], as well as minor agents, such as *Brucella* spp. [16], together with associated online databases.

Four web-based MLVA databases have been developed in France. The first [17], hosted by Université Paris Sud in Orsay, and used worldwide, started in 2002. The third version was released in 2007 and a fourth, which will be able to manage a variety of sequence-based assays in addition to MLVA, is currently under development. The second database [13], developed by the Robert Picqué Military Hospital, was released in 2007. Importantly these two websites allow external users to create their own database, with user-defined species, set of loci, etc., independently of the hosting institution. The resulting databases can be shared within a community or even made publicly accessible. The other two MLVA databases were developed by the Institut Pasteur in Paris [18] and Guadeloupe [19]; the latter is dedicated to *M. tuberculosis*. A list of websites hosting MLVA genotyping databases for a number of pathogens is maintained at the genomes and polymorphisms website [8].

A number of French national or regional reference laboratories are now shifting to, or at least evaluating MLVA as a first-line typing tool: this is the case, for instance, for the *A. baumanii*, *Burkholderia*, *L. pneumophila* and *S. aureus* reference laboratories.

The following section focuses on the use of MLVA for enteropathogenic bacteria genotyping in France.

Use of MLVA for enteric pathogens

In France, laboratory-based approaches are a key component of monitoring strategies for enteric pathogens, as a voluntary laboratory-based network of clinical and veterinary laboratories send bacterial isolates to the National Reference Centre (NRC), which performs serotyping analysis and runs weekly outbreak detection algorithms [20]. The basic information currently provided by French laboratories to public health surveillance is the serotype of isolates; however, the discriminatory capacity is limited. Only a few serotypes are highly prevalent worldwide: Typhimurium and Enteritidis for *Salmonella*, *sonnei* for *Shigella* and O157 for enterohaemorrhagic *Escherichia coli* (EHEC). Differentiation between isolates of the most common serotypes requires the use of subtyping methods: in France, this is carried out by the national reference centres or national veterinary laboratories.

Standardised MLVA schemes for two *Salmonella* serotypes, Typhimurium and Enteritidis, have been used in France since 2005 and 2006, respectively [2,7]. For *S.* Typhimurium and its monophasic variant, the most common *Salmonella* serotypes identified in France from humans and non-humans, the reference laboratories use the widely accepted MLVA nomenclature [2]. Due to a high number of Typhimurium and 4,[5],12:i:- strains collected from humans by the French National Reference Centre annually – around 4,000 and 1,000 respectively [21] – MLVA is exclusively used for outbreak investigations to complement primarily molecular subtyping, i.e. PFGE or clustered regularly interspaced short palindromic repeats (CRISPR) analysis. MLVA is particularly performed to compare strains with those notified from an outbreak in other European countries or to discriminate among clonal isolates indistinguishable by PFGE or CRISPR analysis, such as those belonging to the multidrug-resistant DT104 serotype Typhimurium population or to the egg-related PT4 Enteritidis. A total of 1,252 *Salmonella* clinical isolates were tested by MLVA in France from 2005 to 2011. Of 879 *S.* Typhimurium strains, there were 380 profiles; of 373 monophasic variant strains, there were 40 profiles, suggesting that the 4,[5],12:i:-clone has emerged recently.

Shigella sonnei is a monomorphic organism and therefore requires a highly discriminative sequence-based method for investigations. In France, *S. sonnei* outbreaks have been described and some have been investigated using an eight-loci MLVA scheme with a good Simpson diversity value, as previously described [22].

For *E. coli* O157, MLVA is not performed routinely, as PFGE is sufficient for tracking outbreaks, but it could be used for characterisation of an epidemic clone.

Germany

At the National Reference Laboratory for the Analysis and Testing of Zoonoses (*Salmonella*) in Berlin, MLVA is applied for outbreak studies involving S.Typhimurium, monophasic S.Typhimurium and S. Enteritidis. For *S.* Typhimurium, the standardised protocol [1,2] is used and for S. Enteritidis, the method published by Malorny

Table 2

MLVA analysis of *Salmonella enterica* subspecies *enterica* serovar Typhimurium phage type DT104 strains, Germany, January–April 2010 (n=44)

Row number	Source of *S.* Typhimurium DT104 isolates	Month of isolation	Antiobiotic resistance[a]	Allele string of VNTR loci
Isolates from the sentinel region				
1	29 cases	Mar–Apr	A, C, T, S, Su, Nal	3-14- 9-19-311
2	1 case	Mar	A, C, T, S, Su, Nal	3-14-10-19-311
3	1 case	Mar	A, C, T, S, Su, Nal	3-14- 9-20-311
4	1 isolate (raw sausage)	Mar	A, C, T, S, Su, Nal	3-14- 9-19-311
5	1 isolate (pork)	Mar	A, C, T, S, Su	3-13-14-16-111
6	1 isolate (pork)	Mar	A, C, T, S, Su	3-14-14-16-111
7	1 case	Feb	A, C, T, S, Su	3-14- 3-20-311
8	1 case	Jan	A, C, T, S, Su	3-13- 5-12-311
9	2 cases	Jan	A, C, T, S, Su	3-17-12-16-111
Phenotypically similar isolates from geographically distant regions of Germany				
10	1 case	Jan	A, C, T, S, Su, Nal	3-16- 3-23-311
11	1 case	Jan	A, C, T, S, Su, Nal	3-10-20-12-311
12	1 case	Feb	A, C, T, S, Su, Nal	3-14-18-23-311
13	1 case	Mar	A, C, T, S, Su, Nal	3-14- 9-19-311
14	1 case	Mar	A, C, T, S, Su	3-13- 5-12-311
15	1 case	Apr	A, C, T, S, Su, Nal	3-12-14-16-311

A: ampicillin; C:chloramphenicol; MLVA: multilocus variable-number tandem repeat analysis; Nal: nalidixic acid; S: streptomycin; Su: sulphonamide; T: (oxy)tetracycline; VNTR: variable-number tandem repeat.

[a] Based on antibiogram results. Antibiotic susceptibility testing was performed by broth microdilution method [24]. Breakpoints for interpretation of minimum inhibitory concentration (MIC) values were derived from the European Committee on Antimicrobial Susceptibility Testing (EUCAST) epidemiological cut-off values [25].

et al. [23] is used. The reference laboratory performs about 10 outbreak and tracing studies per year.

S.Typhimurium surveillance in Germany relies initially on phage typing. At the National Reference Center for *Salmonella* and other Enterics in Wernigerode, each year, about 200 to 300 human clinical *S.* Typhimurium isolates from a large sentinel region (five federal states in the middle and west of Germany) are phage typed and kept in a strain collection. Over the past five years, 30% to 10% (decreasing annually) of these isolates were of phage type DT104. However, in March and April 2010, 38 (49%) of all 77 *S.* Typhimurium isolates obtained from this region were of phage type DT104. Strikingly, 34 of these DT104 isolates revealed resistance to nalidixic acid, in contrast to none of the six DT104 isolates from January and February that year. Moreover, all of the 74 *S.* Typhimurium isolates with nine different non-DT104 phage types obtained from the sentinel region between January and April 2010 were susceptible to nalidixic acid. The most obvious explanation for such a substantial increase in the number of *S.* Typhimurium isolates with the phenotypic-character combination of phage type DT104 and nalidixic acid resistance would be a local outbreak. Here we outline hitherto unpublished data on how MLVA was used to identify the outbreak clone.

Searching for a potential source of the infections, regional public health authorities isolated *S.* Typhimurium from several food samples from within the sentinel region; among these were DT104 isolates from pork carcasses and from raw sausages, made in a butcher's shop as a regional delicacy. The DT104 isolates from the carcasses were not resistant to nalidixic acid, but those from the sausages were. We subjected all clinical and food DT104 isolates obtained from January to April 2010 from the sentinel region to MLVA analysis. In addition, we included several phenotypically similar isolates from sporadic cases obtained during the same period from geographically distant regions of Germany. The MLVA results are summarised in Table 2.

Identical MLVA patterns were observed among the majority of clinical *S.* Typhimurium DT104 isolates resistant to nalidixic acid and the raw-sausage isolates (Table 2, rows 1 and 4). It is interesting to note that in two phenotypically indistinguishable isolates there were single locus allelic variants (Table 2, rows 2 and

3), affecting the loci STTR6 and STTR10, respectively. In each case, one locus differed by the presence of one additional repeat unit at the respective VNTR site, compared with the outbreak strain MLVA pattern (Table 2, row 1). Therefore, these loci might well be hypervariable, i.e. drifting towards diversity even within a given outbreak. Attention must be paid to such possible hypervariability, particularly when attempting to use MLVA for long-term surveillance. The phenotypically indistinguishable but spatially and/or temporally independent *S.* Typhimurium isolates, however, (Table 2, rows 5 to 15) were clearly distinguishable by the MLVA approach used.

Ireland

MLVA is used in Ireland for *Salmonella* subtyping: at the National Reference Laboratory (NRL) for *Salmonella* in County Kildare, its use is related to food, animal feed and animal health; MLVA subtyping for public health is carried out at the National *Salmonella* Reference Laboratory, Galway. All *Salmonella* strains isolated from official and food business operator control programmes are submitted to the NRL for typing and this provides an accurate picture of the diversity of *Salmonella* strains circulating in Ireland. Although S. Enteritidis and *S.* Typhimurium are virtually absent in poultry production due to a stamp out policy, *S.* Typhimurium, including the monophasic variant, is frequently isolated largely due to targeted sampling in the pig sector, where the serotype is prevalent. *S.* Typhimurium is also frequently isolated from samples of bovine or equine origin. More extensive information can be found in the 2011 annual report from the NRL for *Salmonella* in food, feed and animal health [26].

The NRL for *Salmonella* uses the standardised MLVA assay [1,2]. This method was initially set up in 2009 using the MegaBACE 1000 but since 2011, it has been based on the ABI 3500 platform. MLVA is applied to ascertain epidemiological linkages between isolates from different sources, e.g. to investigate transmission through the food chain or to prove cross-contamination in specific settings. It has also been very useful to characterise strains related to outbreaks. One such outbreak began in the autumn of 2009 and continued into 2010: the outbreak strain was clearly identified by its distinctive phage type, DT8, and by being fully susceptible to antimicrobials [27]. The MLVA pattern was observed to be either 2-9-NA-12-0212 or 2-10-NA-12-0212. Reported consumption of or exposure to duck eggs explained 70% of cases. Trace-back investigations identified *S.* Typhimurium DT8 with indistinguishable MLVA types from several egg-laying duck flocks. Controls have been introduced in duck egg production units and testing has continued, which has demonstrated *S.* Typhimurium DT8 in over 30 sites (unpublished data).

Another example of the use of MLVA is the retrospective study that was conducted to characterise porcine *S.* Typhimurium isolates recovered from different points in the food chain, from farms to meat processing establishments [28]. It compared the effectiveness of MLVA, phage typing and antimicrobial susceptibility testing in discriminating isolates for epidemiological purposes. From 301 isolates, 154 MLVA patterns were obtained, compared with 19 phage types and 38 antimicrobial resistance patterns. MLVA was particularly useful for discriminating between isolates of the same or similar phage type, e.g. DT104 and DT104b, or isolates that were untypable or in the category of 'reacts with phage but does not conform to a recognised phage type' (RDNC) by phage typing. Cluster analysis of MLVA profiles demonstrated two major clusters (I and II), which had a clear association with particular phage types: cluster I isolates were associated with phage types DT104, U302 and DT120; cluster II with DT193 and U288. The study showed that MLVA was highly discriminatory and permitted the identification of indistinguishable profiles among isolates obtained at different points of the pork food chain.

Italy

Brucellosis is an important zoonosis caused by members of the genus *Brucella*, which is endemic in the south of Italy, and in particular in Sicily. In addition, *Brucella* spp. represent potential biological warfare agents. Since 1995, the availability of whole genome sequences has enhanced the development of multilocus VNTR-based typing approaches such as MLVA. In 2006, a scheme called MLVA-15 – based on a subset of 15 loci that comprises eight markers with good species-identification capability and seven with higher discriminatory power – was published [29], followed by MLVA-16, a slight modification of MLVA-15 [16]. The MLVA band profiles obtained can be resolved by techniques such as agarose gel electrophoresis, microfluidics technology and DNA sequencing. The Dipartimento Sanità Pubblica Veterinaria e Sicurezza Alimentare (Department of Veterinary Public Health and Food Safety) of the national public health institute, Istituto Superiore di Sanità, performs MLVA-15 by direct sequencing of the PCR fragments [30]. The molecular biology section, Centro Studi e Ricerche di Sanità e Veterinaria (CSRSV), of the Italian Army developed a high-throughput system of MLVA-15 and -16 typing for *Brucella* spp. using 'lab-on-a-chip' technology [31,32]. Furthermore, the CSRSV and the National Reference Center for Brucellosis in Italy, Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise Giuseppe Caporale, are developing a new high-throughput *Brucella* genotyping system based on capillary gel electrophoresis.

Human anthrax is currently rare in Italy, the last case was reported in 2006 [33], while for fatal cases, only 27 were reported from 1969 to 1997 [34,35]. Animal cases are mainly located in central and southern Italy, where anthrax is still enzootic, as in other Mediterranean areas. The Centro Studi e Ricerche di Sanità e Veterinaria (CSRSV) has developed the most discriminatory MLVA-based method for subtyping *Bacillus*

anthracis [15], worldwideadopted, based on the analysis of 25 VNTR markers on an automated platform. In 2006, 73 Italian B. anthracis samples were typed by this method, showing that most of the Italian strains were located in the A1.a group, but some strains isolated in northern Italy belonged to B or D groups. This result was an important novelty compared with previous data published in 2005 [36], in which MLVA analysis of 64 Italian isolates revealed that the majority of strains (63/64) belonged to the genetic cluster A1.a, while one isolate was associated with the A3.b cluster. A more recent report (2011) confirmed that in northern Italy strains belonging to the B group could be isolated [37]. This B lineage is present in Italy, the French Alps, Germany and Croatia, so it could be assumed that B genotypes persist in livestock in the French and Italian Alps.

Clostridium botulinum, the etiological agent of botulism, caused in Italy between 2006 and 2011 about 137 botulism cases, one of the highest prevalences in Europe [38]. The reference centre for botulism in Italy is the Centro Nazionale di Riferimento per il botulismo (CNRB), which is part of the Istituto Superiore di Sanità. CNRB maintains a collection of more than 400 *Clostridium botulinum* strains, characterised by phenotypic as well as and genotypic approaches. At CSRSV, a MLVA-15 research project has been developed for *C. botulinum* in collaboration with laboratories of the other countries participating in the European Biodefence Laboratory Network (EBLN). Strains were provided mainly by the CNRB and also by other EBLN institutions. This MLVA scheme improved the discriminatory power compared with the previous MLVA-10 scheme for *C. botulinum* [39]. The analysis was extended to B and F toxin serotype strains, in addition to A serotype strains: five newly characterised MLVA loci were added to the previous 10-MLVA scheme and new groups were described. To date, MLVA data have been obtained for about 300 international *C. botulinum* strains, whereas profiles from 79 strains across Europe have been published [40].

The Netherlands

In the Netherlands, MLVA is used to characterise several pathogenic bacterial species, in research settings and for surveillance purposes. The molecular typing profiles are used to study transmission routes and assess sources of infection and also to assess the impact of human intervention, such as vaccination and use of antibiotics on the composition of bacterial populations. MLVA schemes have been developed and used by several groups outside the National Institute for Public Health and the Environment (RIVM) for the typing of several pathogens, e.g. vancomycin-resistant enterococci [41] and gonococci [42].

Within RIVM, several MLVA schemes have been developed, which are currently used for surveillance of, for example, meticillin-resistant *S. aureus* (MRSA), *S. pneumoniae*, *Bordetella pertussis*, *Haemophilus influenzae* serotype b and *Neisseria meningitidis*. In addition, the national reference laboratory for tuberculosis, located within RIVM, uses the MIRU typing assay (24-loci MLVA) for *M. tuberculosis*. The MLVA schemes developed at RIVM and a typing tool for these pathogens are maintained at RIVM [43]. The typing tool allows interrogation of a MLVA-type table: by typing in an MLVA allelic profile, it will report both the MLVA type and MLVA complex. The tool can be set to report the exact and closest matching profiles.

MLVA of MRSA is by far the most intensely used MLVA scheme in RIVM. By May 2012, the MRSA MLVA database contained MLVA profiles of nearly 29,000 isolates and 3,351 different profiles and 28 MLVA complexes were recognised among these isolates. For MRSA, virtually all isolates are sent to RIVM for molecular typing as part of the national MRSA surveillance. The S. pneumoniae database is the second largest MLVA database at RIVM. Although smaller, it still contains profiles of approximately 4,000 isolates.

In all MLVA schemes used in RIVM, assessment of the number of repeats in each locus is performed by sizing of the fluorescently labelled PCR products on an automated DNA sequencer. Each unique MLVA profile is given a MLVA type designation, e.g. MT21, and profiles are used for clustering and assignment of MLVA complexes. The use of fluorescent labels also allows for the simultaneous MLVA and detection of particular genes. This was used in the MRSA MLVA protocol, in which primer sets were included to detect the *mecA* and *lukF* genes.

Although separation of the PCR products is performed on a DNA sequencer, standardisation may pose a problem for MLVA. Differences may be caused by the use of different sequencers, buffers, etc. In order to compensate for these effects, RIVM supplies calibration sets (shipping costs only) that contain mixtures of PCR products of all known alleles for a particular scheme. Such a calibration set will reveal the positions to which the alleles will migrate on the user's sequencer and will help to define the correct bin positions.

Norway

In Norway, the Norwegian Institute of Public Health (NIPH) is the primary facility for nationwide surveillance of food-borne infections. MLVA is used extensively as the primary routine genotyping tool for a number of enteropathogenic bacteria with the exception of *Campylobacter* spp. (for which other methods are applied), giving the NIPH an up-to-date overview of the spread and introduction of these pathogens in Norway. NIPH genotypes and maintains databases for *E. coli*, *S.* Typhimurium, *Shigella* spp. *Yersinia enterocolitica* and *Listeria monocytogenes*. For typing *E. coli*, three different protocols are in use: two designed for *E. coli* O157:H7 and sorbitol-fermenting O157:H- strains (unpublished), as well as a generic MLVA assay able to genotype all serotypes of *E. coli* using 10 loci [44]. In

2011, 509 *E. coli* isolates were routinely typed using the generic *E. coli* MLVA assay, giving rise to 348 distinct genotypes, with no major outbreaks detected.

The MLVA assays have proven to be highly valuable in strain surveillance and outbreak detection in Norway. It is the speed and resolution of MLVA in particular that has made it the primary genotyping method at NIPH. MLVA data are further coupled with data from virulence-gene assays, phylogenetic-group typing, antibiotic resistance data (if available) or other typing methods such as binary-gene typing or single-nucleotide polymorphism (SNP)-typing to describe the pathogens in detail. In case of a suspected outbreak, other complementary data (e.g. epidemiological) are added as well. A recent review of MLVA typing at NIPH was recently published [45]. Other institutions in Norway have also published MLVA assays: the University of Bergen has published the first MLVA method for typing the fish pathogen *Francisella noatunensis* [46] and the Norwegian Defence Research Establishment (NDRE) has developed and evaluated an MLVA assay for *Vibrio cholerae*, which proved to be both fast (within 3–5 hours) and highly discriminatory [47]. The Norwegian University of Science and Technology has developed and applied an MLVA assay for *Streptococcus agalactiae* with promising results: a five-locus MLVA assay was considered to resolve a strain collection of 126 *S. agalactiae* strains considerably better than multilocus sequence typing (MLST) and with less workload [48].

Sweden

The ease of standardisation and portability of data makes MLVA particularly useful for molecular epidemiology of zoonotic disease agents, where close collaboration between human and animal health agencies is necessary. For example, all primary isolates of *S.* Typhimurium and monophasic *S.* Typhimurium 4,[5],12:i:- found in animals and animal feed are routinely typed at the Swedish National Veterinary Institute (SVA), using the protocol recommended by the European Centre for Disease Prevention and Control (ECDC) [1, 2]. The same method is used for all clinical isolates at the Swedish Institute for Communicable Disease Control (SMI) and data are exchanged continuously to facilitate source attribution and outbreak investigation. The comparability of typing data is ensured by standardised nomenclature and analysis of an external panel of calibration strains [2] at both laboratories.

A similar SMI/SVA collaboration is active for verotoxin-producing *E. coli* (VTEC) O157:H7, using a slightly modified version of the Centers for Disease Control and Prevention protocol developed by Hyytiä-Trees et al. [49]. At SVA, this method has recently been shown to offer comparable performance to PFGE typing for cattle isolates [50], while being substantially faster and less laborious. An ongoing research project is comparing clinical isolate profiles generated at SMI to those from isolates from periodical nationwide slaughterhouse prevalence studies on cattle and from sheep isolates. Again, analysis of a panel of isolates with sequenced loci was necessary to achieve harmonisation between laboratories: in this case, a certain amount of in-house optimisation was also necessary to avoid false negatives due to multiplex PCR competition.

The MLVA for *Coxiella burnetii* at SVA is based on the method by Arricau-Bouvery et al. [51]. In recent years, C. burnetii has been found on several farms in Sweden and by using this method, strains that are prevalent in the country during normal conditions as well as during an outbreak can identified. An advantage of this method is that culturing is not required, which is time consuming and laborious for a biosaftey level (BSL) 3 agent. This method also makes it easier for international collaboration, since there is no need to send live bacteria between countries. For instance, C. burnetii cattle isolate DNA sent to the SVA by a European partner for an epidemiological study is currently being analysed.

In Sweden, there is an increasing trend of pathogenic and non-pathogenic *Enterobacteriaceae* producing extended-spectrum beta-lactamases (ESBL) and plasmid-mediated AmpC (pAmpC) in veterinary settings and food-producing animals. However, compared with the rest of Europe, the problem in Sweden is still very limited, with the exception of the high occurrence of pAmpC and ESBL producing *E. coli* in broilers [52]. SVA is therefore planning to use the extended Lindstedt et al. MLVA protocol [44] to study the genetic relatedness of ESBL- and pAmpC-producing *E. coli* among Swedish broilers, including imported breeding stocks, over time and through the production chain. Collaboration between SVA, SMI and the National Food Agency to compare ESBL-/pAmpC-producing *E. coli* of human, animal and food origin is also in the start-up phase. Furthermore, there are also plans to apply the protocol to study possible outbreaks of ESBL-/pAmpC-producing pathogens in veterinary settings. The same method will also be used in an upcoming SVA/SMI collaborative project for typing of non-O157 VTEC.

Conclusion

Europe has been very successful in developing and using the MLVA methodology: the amount of research and development into MLVA has been considerable for a large array of organisms (Table 1). The development of the methodology within Europe is dynamic and assay updates are frequently published. The first step towards uniform standardisation at the European Union (EU) level has been taken with the online posting of the standard operating procedure for *S.* Typhimurium MLVA by ECDC [53]. This Euroroundup further shows that MLVA has become an important tool for scientific studies and as an aid in outbreak detection and source tracing in European countries.

As MLVA assays rely on the information gathered by genome sequencing, data available for use in method development, or improving exsisting protocols, is being published frequently. As of 17 December 2012, a total of 2,411 whole bacterial genomes were listed by the National Center for Biotechnology Information (NCBI) [54], where all sequences may be downloaded and examined for VNTR content. Thus, MLVA assay development can be performed regardless of access to in-house sequencing (although this is an advantage).

The nature of MLVA makes it a practical system for rapid sharing and digital storing of results, as can be seen by the online databases that are already operational in Europe. This has been achieved in a relatively short time frame: a *S.* Typhimurium MLVA protocol was first published in 2004 [1] and by September 2011, standardised protocols were available in Europe [53]. In comparison, PFGE was first described in the early 80s and it was not until 2004 that PulseNet Europe was established, using protocols standardised in the United States [55]. The modern methodology associated with MLVA protocols makes MLVA a good candidate for integrated surveillance systems, where numerous types of data relating to, for example, strain genotypes, antibiotic resistance, virulence profiles, geographical information and patient/disease information may be stored, combined and shared with the same ease. What is needed is centralised concerted action at the EU level and it is a positive development that ECDC is now integrating MLVA as part of the European Surveillance System (TESSy) [56]. This is an exciting development and it is hoped that more MLVA protocols will be integrated into TESSy in the future. Incorporation of MLVA will be beneficial in outbreak situations where the speed of data retrieval is paramount for source tracing and actions across international borders to end the outbreak.

References

1. Lindstedt BA, Vardund T, Aas L, Kapperud G. Multiple-locus variable-number tandem-repeats analysis of Salmonella enterica subsp. enterica serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J Microbiol Methods. 2004;59(2):163-72.
2. Larsson JT, Torpdahl M, Petersen RF, Sorensen G, Lindstedt BA, Nielsen EM. Development of a new nomenclature for Salmonella typhimurium multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill. 2009;14(15):pii=19174. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19174
3. Torpdahl M, Sorensen G, Lindstedt BA, Nielsen EM. Tandem repeat analysis for surveillance of human Salmonella Typhimurium infections. Emerg Infect Dis. 2007;13(3):388-95.
4. Bruun T, Sorensen G, Forshell LP, Jensen T, Nygård K, Kapperud G, et al. An outbreak of Salmonella Typhimurium infections in Denmark, Norway and Sweden, 2008. Euro Surveill. 2009;14(10). pii=19147. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19147
5. Petersen RF, Litrup E, Larsson JT, Torpdahl M, Sørensen G, Müller L, et al. Molecular characterization of Salmonella Typhimurium highly successful outbreak strains. Foodborne Pathog Dis. 2011; 8(6):655-61.
6. Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, et al. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of Salmonella enterica serotype Enteritidis. J Clin Microbiol. 2007;45:536-43.
7. Hopkins KL, Peters TM, de Pinna E, Wain J. Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of Salmonella enterica serovar Enteritidis. Euro Surveill. 2011;16(32). pii=19942. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19942
8. GMPS. Genomes and PolyMorphismS. Paris: University Paris-Sud. [Accessed 24 Jan 2013]. Available from: http://minisatellites.u-psud.fr
9. Le Flèche P, Fabre M, Denoeud F, Koeck JL, Vergnaud G. High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing. BMC Microbiol. 2002;2:37.
10. Sobral D, Le Cann P, Gerard A, Jarraud S, Lebeau B, Loisy-Hamon F, et al. High-throughput typing method to identify a non-outbreak-involved Legionella pneumophila strain colonizing the entire water supply system in the town of Rennes, France. Appl Environ Microbiol. 2011;77(19): 6899-907.
11. Sobral D, Mariani-Kurkdjian P, Bingen E, Vu-Thien H, Hormigos K, Lebeau B, et al. A new highly discriminatory multiplex capillary-based MLVA assay as a tool for the epidemiological survey of Pseudomonas aeruginosa in cystic fibrosis patients. Eur J Clin Microbiol Infect Dis. 2012;31(9):2247-56
12. Sobral D, Schwarz S, Bergonier D, Brisabois A, Fessler AT, Gilbert FB, et al. High throughput multiple locus variable number of tandem repeat analysis (MLVA) of Staphylococcus aureus from human, animal and food sources. PLoS One. 2012;7(5): e33967.
13. MLVA bacterial genotyping. Bordeaux: Robert Picqué military hospital. [Accessed 24 Jan 2013]. Available from: http://www.mlva.eu
14. Pourcel C, Minandri F, Hauck Y, D'Arezzo S, Imperi F, Vergnaud G, et al. Identification of variable-number tandem-repeat (VNTR) sequences in Acinetobacter baumannii and interlaboratory validation of an optimized multiple-locus VNTR analysis typing scheme. J Clin Microbiol. 2012;49(2): 539-48.
15. Lista F, Faggioni G, Valjevac S, Ciammaruconi A, Vaissaire J, le Doujet C, et al. Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbiology. 2006;6:33.
16. Al Dahouk S, Flèche PL, Nöckler K, Jacques I, Grayon M, Scholz HC, et al. Evaluation of Brucella MLVA typing for human brucellosis. J Microbiol Methods. 2007;69(1): 137–45.
17. Welcome to MLVAbank. Paris: University Paris-Sud. [Accessed 24 Jan 2013]. Available from: http://mlva.u-psud.fr
18. MLVA-NET. Institut Pasteur MLVA database. Paris: Institut Pasteur. [Accessed 24 Jan 2013]. Available from: http://www.pasteur.fr/mlva
19. SITVIT WEB. Pointe-à-Pitre: Institut Pasteur de Guadeloupe. [Accessed 24 Jan 2013]. Available from: http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/
20. David J, Danan C, Chauvin C, Chazel M, Souillard R, Brisabois A, et al. Structure of the French farm-to-table surveillance system for Salmonella. Revue Méd Vét. 2011;162(10): 489-500.

21. Le Hello S, Brisabois A, Accou-Demartin M, Josse A, Marault M, Francart S, et al. Foodborne outbreak and nonmotile Salmonella enterica variant, France. Emerg Infect Dis. 2012;18(1):132-4.
22. Filliol-Toutain I, Chiou CS, Mammina C, Gerner-Smidt P, Thong KL, Phung DC, et al. Global Distribution of Shigella sonnei Clones. Emerg Infect Dis. 2011;17(10):1910-2.
23. Malorny B, Junker E, Helmuth R. Multi-locus variable-number tandem repeat analysis for outbreak studies of Salmonella enterica serotype Enteritidis. BMC Microbiol. 2008;8:84.
24. The International Organization for Standardization (ISO). ISO 20776-1:2006. Clinical laboratory testing and in vitro diagnostic test systems -- Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices -- Part 1: Reference method for testing the in vitro activity of antimicrobial agents against rapidly growing aerobic bacteria involved in infectious diseases. Geneva: ISO; 2006. Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=41630
25. European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 3.0, 2013. EUCAST: 2013. Available from: http://www.eucast.org
26. Department of Agriculture, Food and the Marine (DAFM), Ireland. The 2011 annual report from the National Reference Laboratory for Salmonella. (Food, Feed and Animal Health). Backweston: DAFM; 2011. Available from: http://www.agriculture.gov.ie/media/migration/animalhealthwelfare/labservice/nrl/NRLSalmonellaAnnualReport2011.pdf
27. Garvey P, McKeown P, on behalf of the Outbreak Control Team. Two new cases linked with nationwide 'duck egg' outbreak of Salmonella Typhimurium DT8. Epi-Insight. 2011;12(4). Available from : http://ndsc.newsweaver.ie/epiinsight/3zgoibihqqw87nh5ab6w5b.
28. Prendergast DM, O'Grady D, Fanning S, Cormican M, Delappe N, Egan J, et al. Application of multiple locus variable number of tandem repeat analysis (MLVA), phage typing and antimicrobial susceptibility testing to subtype Salmonella enterica serovar Typhimurium isolated from pig farms, pork slaughterhouses and meat producing plants in Ireland. Food Microbiol. 2011;28(5):1087-94.
29. Le Flèche P, Jacques I, Grayon M, Al-Dahouk A, Bouchon P, Denoeud F, et al. Evaluation and selection of tandem repeat loci for a Brucella MLVA typing assay. BMC Microbiol. 2006;6:9
30. Marianelli C, Graziani C, Santangelo C, Xibilia MT, Imbriani A, Amato R, et al. Molecular epidemiological and antibiotic susceptibility characterization of Brucella isolates from humans in Sicily, Italy. J Clin Microbiol. 2007;45(9):2923-8
31. De Santis R, Ciammaruconi A, Faggioni G, D'Amelio R, Marianelli C, Lista F. Lab on a chip genotyping for Brucella spp. based on 15-loci multi locus VNTR analysis. BMC Microbiol. 2009;9:66.
32. De Santis R, Ciammaruconi A, Faggioni G, Fillo S, Gentile B, Di Giannatale E, et al. High throughput MLVA-16 typing for Brucella based on the microfluidics technology. BMC Microbiol. 2011;11:60.
33. Kreidl P, Stifter E, Richter A, Aschbachert R, Nienstedt F, Unterhuber H, et al. Anthrax in animals and a farmer in Alto Adige, Italy. Euro Surveill. 2006;11(2): pii=2900. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2900
34. Serraino D, Puro V, Bidoli E, Piselli P, Girardi E, Ippolito G. Anthrax, botulism and tularemia in Italy. Infection. 2003;31(2):128-9.
35. Antrace o carbonchio [Anthrax]. Il sito WEB del Sistema Informatizzato Malattie Infettive (SIMI). [The website of the Italian Computerized System for Infectious Diseases (SIMI)]. Rome: Istituto Superiore di Sanità. [Accessed 24 jan 2013]. Available from: http://www.simi.iss.it/antrace_carbonchio.htm
36. Fasanella A, Van Ert M, Altamura SA, Garofolo G, Buonavoglia C, Leori GJ, et al. Molecular diversity of Bacillus anthracis in Italy. J Clin Microbiol. 2005;43(7):3398-401.
37. Garofolo G, Serrecchia L, Corrò M, Fasanella A. Anthrax phylogenetic structure in Northern Italy. BMC Res Notes. 2011;4:273.
38. European Centre for Disease Prevention and Control (ECDC). Annual Epidemiological Report 2011. Reporting on 2009 surveillance data and 2010 epidemic intelligence data. Stockholm: ECDC; 2011. Available from: http://ecdc.europa.eu/en/publications/Publications/1111_SUR_Annual_Epidemiological_Report_on_Communicable_Diseases_in_Europe.pdf
39. Macdonald, TE, Helma CH, Ticknor LO, Jackson PJ, Okinaka RT, Smith LA, et al. Differentiation of Clostridium botulinum serotype A strains by multiple-locus variable-number tandem-repeat analysis. Appl Environ Microbiol. 2008;74(3):875-82.
40. Fillo S, Giordani F, Anniballi F, Gorgé 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):4252-43.
41. Top J, Banga NM, Hayes R, Willems RJ, Bonten MJ, Hayden MK. Comparison of multiple-locus variable-number tandem repeat analysis and pulsed-field gel electrophoresis in a setting of polyclonal endemicity of vancomycin-resistant Enterococcus faecium. Clin Microbiol Infect. 2008;14(4):363-9.
42. Heymans R, Schouls LM, van der Heide HG, van der Loeff MF, Bruisten SM. Multiple-locus variable-number tandem repeat analysis of Neisseria gonorrhoeae. J. Clin. Microbiol. 2011;49(1):354–63.
43. National Institute for Public Health and the Environment (RIVM). MLVA. Bilthoven: RIVM. . [Accessed 24 jan 2013]. Available from: http://www.mlva.net/
44. Løbersli I, Haugum K, Lindstedt BA. Rapid and high resolution genotyping of all Escherichia coli serotypes using 10 genomic repeat-containing loci. J Microbiol Methods. 2012;88(1):134-9.
45. Lindstedt BA. Genotyping of selected bacterial enteropathogens in Norway. Int J Med Microbiol. 2011;301(8):648-53.
46. Brevik ØJ, Ottem KF, Nylund A. Multiple-locus, variable number of tandem repeat analysis (MLVA) of the fish-pathogen Francisella noatunensis. BMC Vet Res. 2011;7:5.
47. Olsen JS, Aarskaug T, Skogan G, Fykse EM, Ellingsen AB, Blatny JM. Evaluation of a highly discriminating multiplex multi-locus variable-number of tandem-repeats (MLVA) analysis for Vibrio cholerae. J Microbiol Methods. 2009;78(3):271-85.
48. Radtke A, Bruheim T, Afset JE, Bergh K. Multiple-locus variant-repeat assay (MLVA) is a useful tool for molecular epidemiologic analysis of Streptococcus agalactiae strains causing bovine mastitis. Vet Microbiol. 2012;157(3-4):398-404.
49. Hyytiä-Trees E, Smole SC, Fields PA, Swaminathan B, Ribot EM. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing Escherichia coli O157 (STEC O157). Foodborne Pathog Dis. 2006;3(1):118-31.
50. Eriksson E, Soderlund R, Boqvist S, Aspan A. Genotypic characterization to identify markers associated with putative hypervirulence in Swedish Escherichia coli O157:H7 cattle strains. J Appl Microbiol. 2011;110(1):323-32.
51. Arricau-Bouvery N, Hauck Y, Bejaoui A, Frangoulidis D, Bodier CC, Souriau A, et al. Molecular characterization of Coxiella burnetii isolates by infrequent restriction site-PCR and MLVA typing. BMC Microbiol. 2006;6:38.
52. Bengtsson B, Ericsson Unnerstad H, Greko G, Grönlund Andersson U, Landen A, editors. Swedish Veterinary Antimicrobial Resistance Monitoring (SVARM) 2010. Uppsala: The National Veterinary Institute (SVA); 2010. Available from: http://www.sva.se/upload/Redesign2011/Pdf/Om_SVA/publikationer/1/Svarm2010.pdf
53. European Centre for Disease Prevention and Control (ECDC). Laboratory standard operating procedure for MLVA of Salmonella enterica serotype Typhimurium. Stockholm: ECDC; 2011. Available from: http://ecdc.europa.eu/en/publications/Publications/1109_SOP_Salmonella_Typhimurium_MLVA.pdf
54. Genome information by organism. Bethesda: National Center for Biotechnology Information, United States National Library of Medicine. [Accessed 17 Dec 2012]. Available from: http://www.ncbi.nlm.nih.gov/genome/browse/
55. PulseNet Europe. The European Molecular Subtyping Network for Foodborne Disease Surveillance. [Accessed 24 jan 2013]. Available from: http://www.pulsenetinternational.org/networks/Pages/europe.aspx
56. European Centre for Disease Prevention and Control (ECDC). The European Surveillance System (TESSy). Stockholm: ECDC. [Accessed 24 Jan 2013]. Available from: http://ecdc.europa.eu/en/activities/surveillance/tessy/pages/tessy.aspx

Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections

C W Knetsch[1], T D Lawley[2], M P Hensgens[1], J Corver[1], M W Wilcox[3], E J Kuijper (E.J.Kuijper@lumc.nl)[1]
1. Section Experimental Microbiology, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, Netherlands
2. Bacterial Pathogenesis Laboratory, Wellcome Trust Sanger Institute, Hinxton, United Kingdom
3. Microbiology Department, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

Citation style for this article:
Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. Current application and future perspectives of molecular typing methods to study Clostridium difficile infections. Euro Surveill. 2013;18(4):pii=20381. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20381

Article submitted on 25 July 2012 / published on 24 January 2013

Molecular typing is an essential tool to monitor *Clostridium difficile* infections and outbreaks within healthcare facilities. Molecular typing also plays a key role in defining the regional and global changes in circulating *C. difficile* types. The patterns of *C. difficile* types circulating within Europe (and globally) remain poorly understood, although international efforts are under way to understand the spatial and temporal patterns of *C. difficile* types. A complete picture is essential to properly investigate type-specific risk factors for *C. difficile* infections (CDI) and track long-range transmission. Currently, conventional agarose gel-based polymerase chain reaction (PCR) ribotyping is the most common typing method used in Europe to type *C. difficile*. Although this method has proved to be useful to study epidemiology on local, national and European level, efforts are made to replace it with capillary electrophoresis PCR ribotyping to increase pattern recognition, reproducibility and interpretation. However, this method lacks sufficient discriminatory power to study outbreaks and therefore multilocus variable-number tandem repeat analysis (MLVA) has been developed to study transmission between humans, animals and food. Sequence-based methods are increasingly being used for *C. difficile* fingerprinting/typing because of their ability to discriminate between highly related strains, the ease of data interpretation and transferability of data. The first studies using whole-genome single nucleotide polymorphism typing of healthcare-associated *C. difficile* within a clinically relevant timeframe are very promising and, although limited to select facilities because of complex data interpretation and high costs, these approaches will likely become commonly used over the coming years.

Introduction

Clostridium difficile is a gram-positive rod-shaped anaerobic bacterium that is capable of forming spores. Since its discovery as a cause of antibiotic-associated pseudomembranous colitis nearly 30 years ago [1], *C. difficile* has become the major cause of antibiotic-associated diarrhoea. Antibiotics change the protective normal gut flora, which enables *C. difficile* to colonise the colon. Clinical symptoms may range from simple diarrhoea to severe colitis which can result in death [2]. Symptoms are primarily mediated by two virulence factors, toxins A (tcdA) and B (tcdB), which are released in the gut upon colonisation by *C. difficile* [3-5]. In the past decade, the epidemiology of *C. difficile* has changed and a new type emerged: polymerase chain reaction (PCR) ribotype (RT) 027/North American pulsed (NAP)-field type 01. Besides the production of toxins A and B, the binary *C. difficile* transferase toxin A/B (cdtA and cdtB) has probably contributed to the increased virulence of this type in addition to still unknown factors [6]. Major outbreaks due to this strain were reported since 2004, first in Canada followed by North America and Europe [7-10]. In 2008, PCR RT078/NAP07-08 was reported as an emerging strain [11].

To study the epidemiology of *C. difficile*, several molecular typing methods have been introduced. Ideally, a typing method must have sufficient discriminatory power, typeability (the ability to type isolates unambiguously), reproducibility and transportability (the ability to perform the method reproducibly in a fully compatible fashion in different laboratories at different times) and must be relatively easy to perform [12]. In this review, we describe the most commonly used typing methods to characterise *C. difficile*. In addition, we present the latest developments in typing of *C. difficile*. Finally, we discuss the use of typing in surveillance studies, to trace outbreaks and to study strain transmission from the environment to patients.

Historical perspective of *Clostridium difficile* typing

Molecular typing methods can be categorised into two groups, phenotypic and genotypic methods. In the 1980s only phenotypic techniques were available. Serotyping using slide agglutination was commonly used in the mid-1980s. Initially, this assay was capable to differentiate six serogroups [13], later this

Table 1

Performance characteristics of various genotyping methods for *Clostridium difficile*

Method	Target	Discriminatory power	Typeability	Reproducibility	Ease of interpretation	Technical complexity	Transportability
Band-based							
REA	Whole genome	Good	Fair	Fair	Poor	Moderate	Poor
PFGE	Whole genome	Moderate	Fair	Moderate	Fair	Moderate	Moderate
PCR ribotyping	16S–23S ISR	Good	Moderate	Moderate	Moderate	Low	Moderate
Capillary PCR ribotyping	16S–23S ISR	Excellent	Moderate	Good	Good	Moderate	Good
MLVA	Whole genome, tandem repeats	Excellent	Poor	Moderate	Good	Moderate	Moderate
Sequence-based							
MLST 7HG	7 HG	Good	Moderate	Moderate	Excellent	Moderate	Excellent
SNP typing	Whole genome, SNPs	Excellent	Moderate	Moderate	Excellent	High	Good

HG: housekeeping genes; ISR: intergenic spacer region; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis; PCR: polymerase chain reaction; PFGE: pulsed-field gel electrophoresis; REA: restriction endonuclease analysis; SNP: single nucleotide polymorphism.

Table modified from Kuijper et al. [17].

was improved to 15 serogroups [14]. Other commonly used methods in this period were autoradiography polyacrylamide gel electrophoresis (radio PAGE) [15] and immunoblotting using rabbit antiserum prepared from rabbits immunised with four different *C. difficile* strains [16]. Phenotypic assays had low reproducibility, low typeability and insufficient discriminatory power to apply to epidemiological studies [12]. Genotypic techniques with better typeability and discriminatory power replaced phenotypic methods during the 1990s [12]. Genotypic methods are divided into band-based and sequence-based methods. The most commonly used band-based methods were restriction endonuclease analysis (REA), pulsed-field gel electrophoresis (PFGE), capillary or conventional PCR ribotyping and multilocus variable-number tandem repeat analysis (MLVA), whereas the most frequently used sequence-based genotyping method was multilocus sequence typing (MLST). Recently whole genome sequencing (WGS) has emerged as a promising sequence-based technique as it allows the detection of variations between *C. difficile* strains by, for example, single nucleotide polymorphisms (SNPs) analysis. Here we present a brief summary of the current performance and costs of genotyping methods (Table 1 and 2), as a detailed description is beyond our scope and can be found in three other reviews on molecular typing [12,17,18].

Currently used typing methods for *Clostridium difficile*

In Europe PCR ribotyping is presently the most frequently used typing method of *C. difficile*. This method was first applied by Gurtler et al. [21] and exploits the variability of the intergenic spacer region (ISR) between the 16S and 23S ribosomal DNA (rDNA), which is type-dependent. The variability, in combination with multiple copies of rDNA present in the genome, results in various amplicons after PCR amplification. These amplicons are separated by common agarose gel electrophoresis. The obtained banding patterns are referred to as PCR RTs. Two different sets of primers have been developed for typing of *C. difficile* [22,23]. The O'Neill primers described by Stubbs et al. [23] seem to have better discriminatory power than the Bidet primers [24]. The discriminatory power (*D*) of a typing method is its ability to distinguish between unrelated strains, this D-value is based on Simpson's index of diversity [25]. PCR ribotyping is currently capable of identifying more than 400 distinct PCR RTs.

In North-America, PFGE is commonly used. PFGE of *C. difficile* involves digestion of genomic DNA with an infrequent cutting restriction enzyme, for example *Sma*I [26]. PFGE allows separation of large DNA fragments which is not possible with conventional agarose gel electrophoresis. The obtained DNA fragments are separated using agarose gel electrophoresis with an electric field orientation repeatedly switching in three different directions (pulsed-field); one direction is through the central axis of the gel, whereas the other two are at an angle of 60 degrees on either side. The pulse time of the direction is linearly increased during the run so that progressively larger fragments are able to migrate forward through the gel, resulting into separation based on fragment size. The obtained banding patterns are referred to as NAP-field types. Unfortunately, standardisation of protocols and validation of PFGE for *C. difficile* have never progressed as they did for other food-borne pathogens on PulseNet at the United States (US) Centers for Disease Control and Prevention (CDC) [27].

Table 2

Techniques, time and costs associated with various genotyping methods for *Clostridium difficile*

Genotyping method	Techniques	Turnaround time (post-culture)	Hands-on time (post-culture)	Costs	
				Equipment[a]	Per test[b]
REA	DI, ER, GE	2 days	2 hours	Low	Low
PFGE	DI, ER, GE	2–4 days	6 hours	Moderate	Low
PCR ribotyping	DI, PCR, GE	1–1.5 days	2 hours	Low/ moderate	Low
Capillary ribotyping	DI, PCR, CE	1 day	2 hours	Moderate/ high	Low
MLVA	DI, PCR, CE	2 days	8 hours	Moderate/ high	Low/ moderate
MLST	DI, PCR, PPP, SE	4 days	8 hours	Moderate/ high	Moderate
SNP typing	DI, LP, TA, SE	5 days[c]	3 days[d]	High	High

CE: capillary electrophoresis; DI: DNA isolation; ER: enzyme restriction; GE: gel electrophoresis; LP: library preparation; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis; PCR: polymerase chain reaction; PFGE: pulsed-field gel electrophoresis; PPP: PCR product purification; REA: restriction endonuclease analysis; SE: sequencing; SNP: single nucleotide polymorphism; TA: template amplification.

[a] Cost index for the equipment set-up: low < EUR 10,000 < moderate < EUR 100,000 < high.

[b] Cost index per test for materials: low < EUR 10 < moderate < EUR 100 < high.

[c] This estimated turnaround time is based on using Illumina Miseq benchtop sequencing [19].

[d] The hands-on time was determined by turnaround time substracted with the average runtime of the Illumina Miseq benchtop sequencer [20].

It has been reported that PFGE displays better discriminatory power than PCR ribotyping with D-values of 0.843 and 0.688, respectively [18]. In contrast, preliminary results of a study comparing different typing techniques on 39 of the most frequently found PCR RTs in Europe demonstrate that only 16 NAP-field types were obtained of 39 PCR RTs (personal communications, M Mulvey and D McCannel, 2011). A common concern with all band-based typing methods is the difficult interpretation of DNA banding patterns, especially when a DNA banding pattern differs marginally from the reference patterns. Consequently, appropriate definitions are required to identify new types with both PFGE and PCR ribotyping. In Europe, the Cardiff collection of Jon Brazier and Val Hall serves as a reference collection and new PCR RTs are always validated using this database. Currently, a clinical collection of 20 different *C. difficile* PCR RTs (European Centre for Disease Prevention and Control (ECDC)-Brazier collection) isolated from various European countries is available to distribute among all reference laboratories in Europe who participate in the European *C. difficile* infection study network (ECDISnet) [28]. The usage of two different standard typing methods in Europe and America has resulted into different nomenclatures, making interlaboratory exchange of data difficult. Already in 1994 Brazier et al. [29] emphasised the need for a unified nomenclature.

In 2004, MLST was introduced to study the population structure and global epidemiology of *C. difficile* [30]. This sequence-based typing method relies on sequencing of DNA fragments approximately ranging between 300 and 500 bp representing seven housekeeping genes (MLST 7HG). Sequence variants for each housekeeping gene are assigned with a distinct allele number and the combination of seven allele numbers (allelic profile) provides a sequence type (ST). MLST generates high-throughput sequence data that can be uploaded from laboratories worldwide to a common web database [31]. This facilitates ST calling as well as studying the population structure and global epidemiology of *C. difficile*. Two different typing schemes have been proposed in literature to characterise *C. difficile* isolates [30,32]. Both typing schemes consist of seven housekeeping genes of which three are shared (triosephosphate isomerase (*tpi*), recombinase A (*recA*) and superoxide dismutase A (*soda*). In contrast to the scheme published by Griffiths et al. [32], the MLST scheme described by Lemee et al. [30] was not widely adopted. This can be partially explained by the presence of a null allele on the D-alanine--D-alanine ligase (*ddl*) locus of the Lemee scheme which failed to amplify in certain strains [32]. Recently, this locus in the Lemee scheme was replaced by the groEL gene [33].

It has been reported that the discriminatory power of MLST and PCR ribotyping is comparable [18,32]. For studying outbreaks at a local level, a typing method should have higher discriminatory power than PCR ribotyping and MLST. For instance an increase in incidence of a PCR RT or MLST ST in a hospital can provide us with a clue for an outbreak and is useful data for monitoring changes in type prevalence rates, but does not necessarily proves clonal spread of one strain.

MLST is an appropriate tool for studying the phylogeny of *C. difficile*. Compared to a band-based typing method, such as PCR ribotyping, MLST is less vulnerable to recombination events. Recombination in a housekeeping gene would change the allelic profile on a single locus only. Even though the consequence would be a change of ST, this new ST would still be closely related to the original ST maintaining the phylogenetic link. Recombination of repeats present in the ISR between the 16S and 23S rDNA [34] might lead to the formation of a novel PCR RT without a clear phylogenetic link. However, the rate at which these recombination events occur and the predisposing factors are unknown. Phylogeny reconstruction with MLST revealed that *C. difficile* diversified into at least five well separated lineages during evolution [32,35,36] and possibly a sixth monophyletic lineage [37]. The majority of STs were assigned to lineage 1 with no major subdivisions (Figure 1), but this result could be due to an unfortunate choice of housekeeping genes. Changing the housekeeping genes or adding housekeeping genes to the current MLST scheme might provide a better resolution of lineage 1.

A major advantage of sequence-based typing methods like MLST is the ease of interpretation of the generated data. Sequence data are unambiguous and therefore objective, highly reproducible and easily exchangeable between laboratories. Moreover, many laboratories have submitted their sequences to a freely accessible *C. difficile* MLST database [31]. Currently (last updated: 21 Nov 2012), 176 different STs have been identified. A practical disadvantage of MLST remains the relatively high cost of sequencing multiple targets, which could partially explain why MLST has not replaced conventional PCR ribotyping in many European laboratories.

MLVA is a highly discriminatory molecular typing method that has been introduced to study outbreaks and identify routes of transmission between patients and hospitals [11,38–42]. MLVA relies on the amplification of short tandem repeats that vary in size and are dispersed throughout the genome. The obtained amplicons are separated with capillary electrophoresis followed by automated fragment analysis. Initially, two different typing schemes were published which both contain seven loci of which four are identical [41,42]. Each of the seven loci is designated with a number that corresponds to the sum of repeats present on that locus. A minimum spanning tree (MST) can be constructed, in which the summed tandem repeat difference (STRD) is used as a measure of genetic difference (Figure 2). Clonal clusters are defined by an STRD of ≤2, and genetically related clusters are defined by an STRD of ≤10 [11,41]. Broukhanski et al. [43] observed that two MLVA loci (F3 and H9) were invariable, indicating that loci F3 and H9 did not contribute to the discriminatory power. In addition, Bakker et al. [44] reported that MLVA locus A6 is a null allele in PCR RT078 and that for several other loci the PCR settings had to be optimised for PCR RT078. Invariance of MLVA loci requires optimisation and validation of MLVA for individual PCR RTs. Currently, MLVA has been implemented as useful typing method to investigate *C. difficile* 027 outbreaks in the Netherlands, France and the United Kingdom (UK) [38,45,46]. In England, *C. difficile* infection (CDI) cases that are potentially linked, i.e. caused by isolates that share the same PCR RT and which are related in time and place, are investigated using MLVA. Notably, almost half of such presumed clusters are shown actually either to consist of unrelated isolates or a mixture of related and distinct strains [46].

Recent developments in typing of *Clostridium difficile*

Variant multilocus variable-number tandem repeat analysis typing schemes

Recently, a modified MLVA (mMLVA) was developed, combining MLVA with PCR detection of several toxin genes (tcdA and tcdB, cdtB; and deletions in the toxin C gene (tcdC)) [37]. In addition, the number of MLVA loci was restricted to five excluding the invariable loci F3 and H9. Although the combination with toxin gene detection can be informative, it is not yet possible to correlate these data with specific *C. difficile* types, like PCR RT027/NAP01. This is partially because the presence of binary toxin genes combined with the 18 bp tcdC deletion is not restricted to PCR RT027 strains [37,47].

In a study by Manzoor et al. [48] the number of MLVA loci was increased to 15. This extended MLVA (eMLVA) scheme was able to discriminate clinically significant clusters while maintaining a good concordance with PCR ribotyping. Typing schemes containing only seven loci showed in contrast poor association with PCR ribotyping [41,42]. These seven loci schemes can only be used as a subtyping method together with PCR ribotyping, whereas the extended MLVA can potentially replace both. It should be noted, however, that increasing the number of loci makes the method more laborious and increases the difficulty of data interpretation.

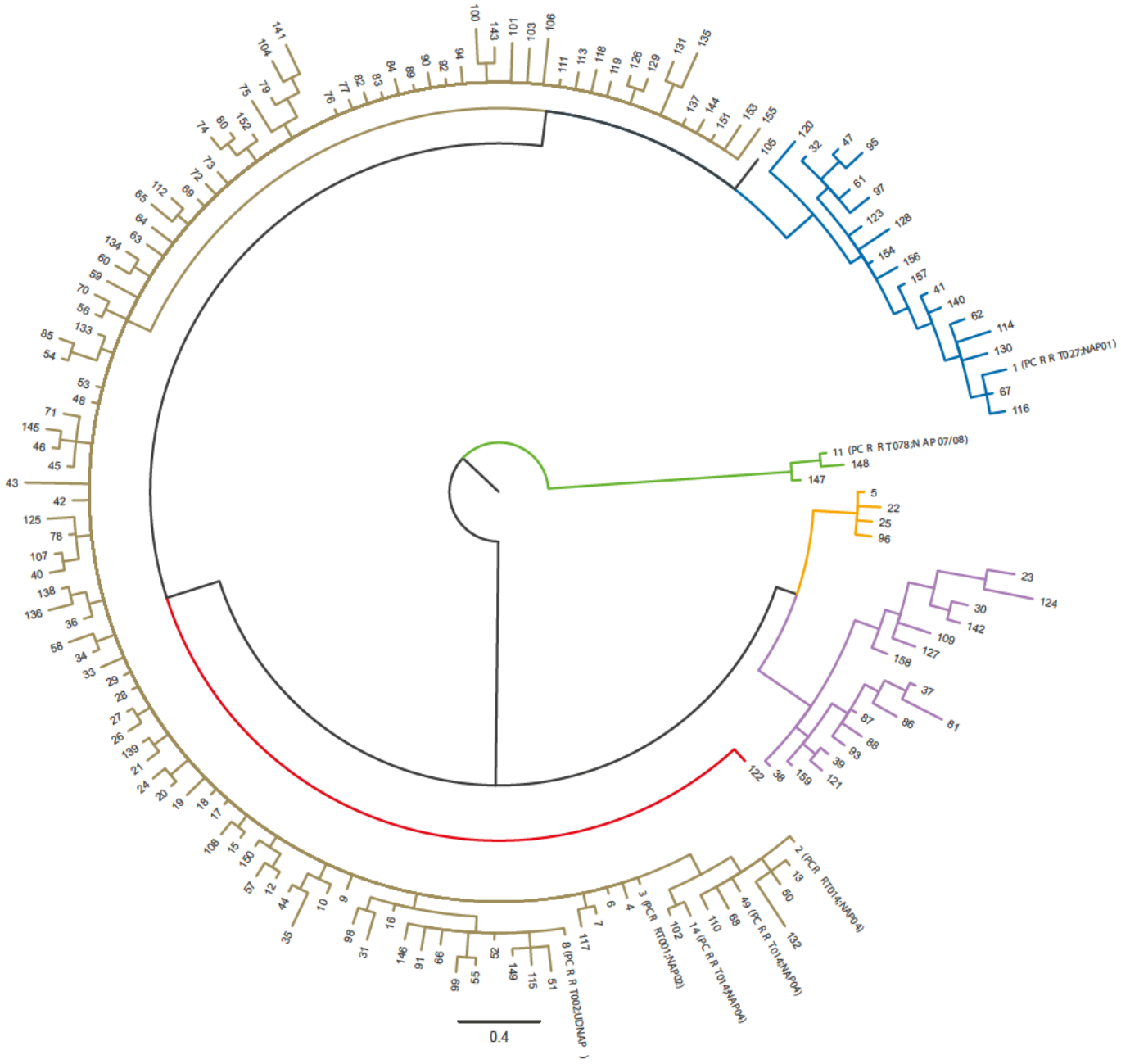
Wei et al. [49] screened 40 MLVA loci for developing an MLVA typing scheme that has a good concordance with PCR ribotyping and provides satisfactory data for studying outbreaks. From this study, it was concluded that typing schemes consisting of MLVA loci with low allelic diversity maintained a high correlation with PCR ribotyping, whereas typing schemes using MLVA loci with high allelic diversity were required to study outbreaks. To fulfil both purposes two different typing schemes were proposed comprising 10 loci with limited allelic diversity and four loci with highly variable allelic diversity.

Capillary polymerase chain reaction ribotyping

Although PCR ribotyping has become widely used in many European laboratories for *C. difficile* surveillance, issues with pattern interpretation and limited access to a well standardised database are

Figure 1

Phylogenetic structure of *Clostridium difficile* strains

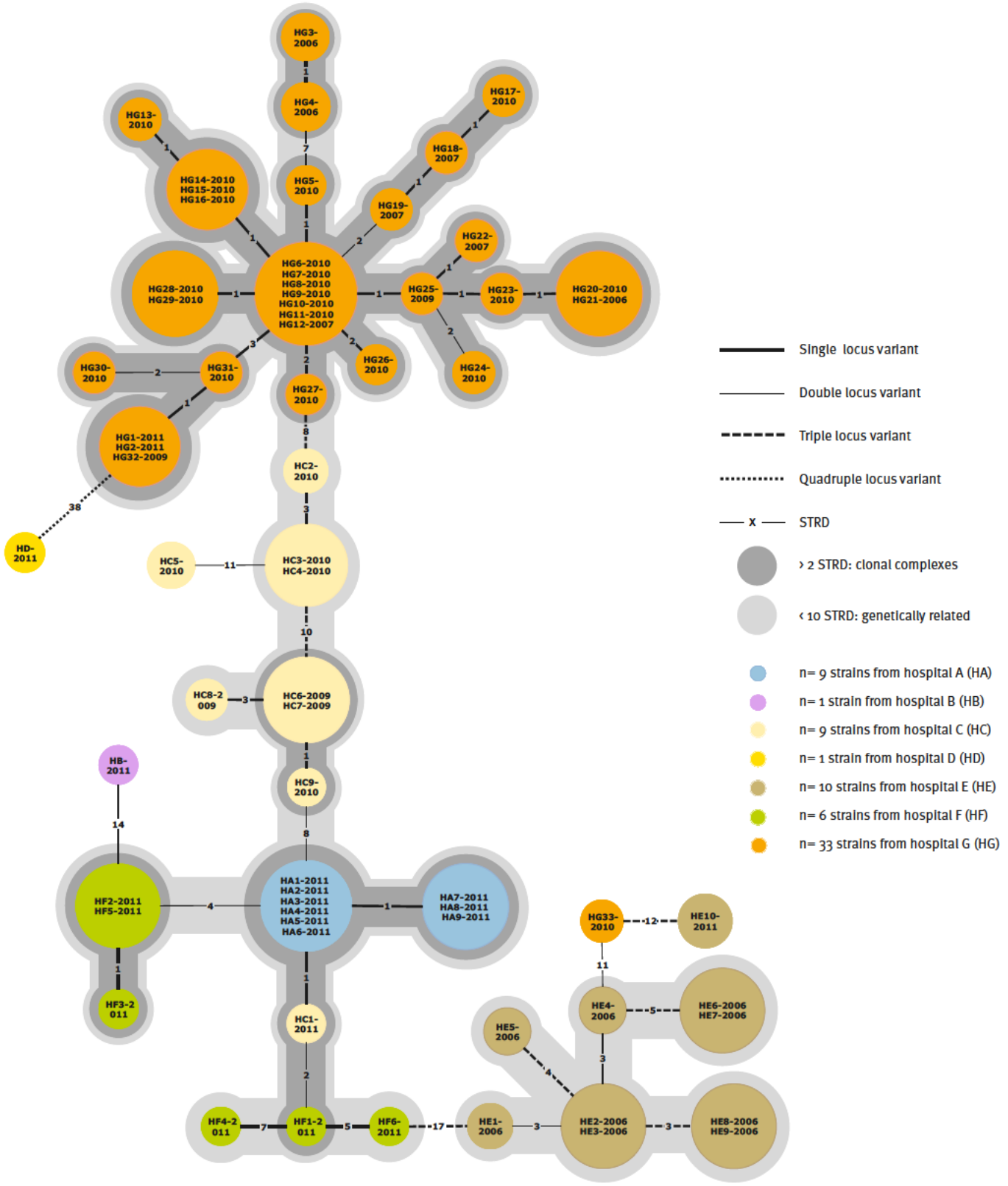


NAP: North American pulsed-field; PCR RTs: polymerase chain reaction ribotypes; UDNAP: undefined NAP field type.

The figure is modified from Knetsch et al. [37]. The phylogenetic tree (radial tree layout) was constructed using a bayesian posterior probability method based on the alignment of concatenated DNA sequences of seven housekeeping gene loci. Six major lineages are shown in colour. The PCR RTs and NAP field types of the five most frequently PCR RTs in Europe are shown between brackets and in bold.

Figure 2

Minimum spanning tree illustrating distinct local *Clostridium difficile* outbreaks



STRD: summed tandem repeat difference.

Multilocus variable-number tandem repeat analysis (MLVA) was used to recognise three different large local outbreaks in hospital G (orange), hospital A (blue) and hospital E (brown). Smaller outbreaks are indicated for hospital C (light yellow), hospital F (green) and related isolates from hospital B (purple) and hospital D (dark yellow). Clonal clusters are defined by a STRD of ≤ 2, and genetically related clusters are defined by an STRD of ≤10.

important limitations. The adaptation of PCR ribotyping to high resolution capillary gel electrophoresis (CE) PCR ribotyping has greatly improved pattern reproducibility and interpretation. For instance, using conventional agarose gel-based PCR ribotyping, it is difficult to differentiate types 014 and 020. In contrast, CE-PCR ribotyping can discriminate type 014 and type 020 and distinguish subtypes within type 014 [50]. However, the need for protocol standardisation remains evident. *C. difficile* surveillance laboratories from the CDC in the US, Public Health Agency of Canada (PHAC) in Canada, Leiden University Medical Center (LUMC) in the Netherlands and Leeds Teaching Hospitals NHS Trust in the UK are collaborating to develop and validate a standardised protocol for the DNA extraction, primer sets, PCR cycling conditions, and reference standards for CE-PCR ribotyping. The standardised consensus protocol is tested on a well characterised collection of 70 different PCR RTs [37] distributed to each of the four laboratories. Preliminary results show consistent fingerprints between the laboratories. Peakfile-based analysis is currently being optimised and validated, with a conclusion available by mid-2013.

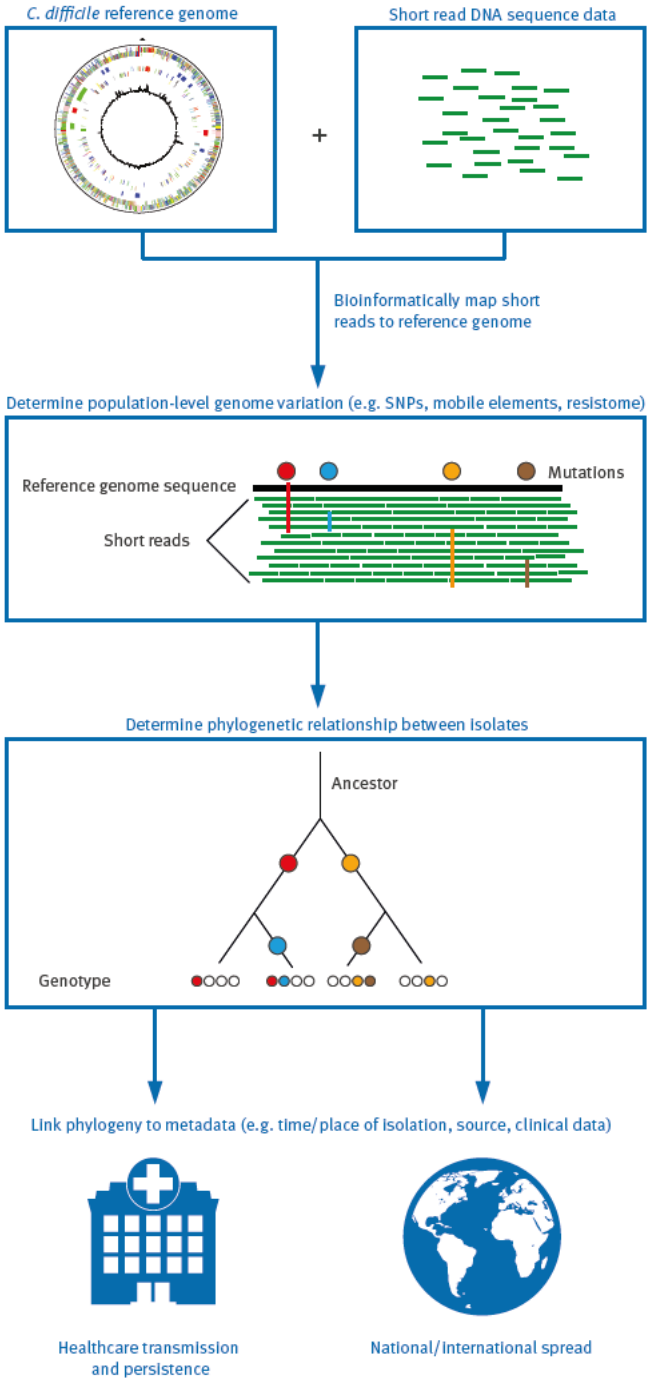
Whole-genome single nucleotide polymorphism typing

High-throughput, WGS of bacterial pathogens has reached a scale and reliability to accurately define the natural history and global population structures of virulent and epidemic lineages [51–55]. Phylogenetic and comparative genome analysis of hundreds (soon to be thousands) of genomes can identify precise genetic changes, often linked to virulence and antibiotic resistance phenotypes, that can quickly inform about the pathogen's biology. Whole genome sequencing can also distinguish between strains at the single nucleotide level, by comparing genomes in terms of single nucleotide polymorphisms, and therefore drastically improves the discriminatory power over conventional genetic typing methods. Thus, WGS has also (i.e. besides phylogeny) practical value for clinical microbiology and public health epidemiology by defining the selective forces that precipitate pathogen emergence and also by tracking transmission events ([56], Figure 3).

WGS approaches represent the ultimate pathogen typing method and, although its use and application remains limited to select facilities, we believe WGS will become a commonly used tool for *C. difficile* surveillance and epidemiology in the coming years. Although the cost of WGS is relatively high compared to traditional typing methods, sequencing costs are falling rapidly [19,57]. In addition, the ability to extrapolate MLST, PFGE, resistance gene, toxin gene sequence and other data from the same test could balance the cost-benefit analysis. Standardised computational pipelines are emerging for *C. difficile* genome data quality control and subsequent downstream analysis associated with informatics, phylogeny and phylogeography (Figure 3). Improved high-quality draft genomes [58] for the most

Figure 3

General sequencing and analysis strategy used to track genomic variants of *Clostridium difficile* at local and global levels



SNPs: single nucleotide polymorphisms.

Genomic DNA derived from *Clostridium difficile* isolates under study are subjected to sequencing with next generation sequencing technologies. Short read data from next generation sequencing platforms are mapped to reference genomes to determine the population level genome variation, such as SNPs, mobile element or other signatures of selection. Isolate sequences of interest are phylogenetically analysed. Combining phylogeny to epidemiological sequence data allows for inferences to be made about pathogen evolution and transmission events at healthcare and global level.

common *C. difficile* variants causing disease in human and animal populations [59] serve as references to map next generation sequence data in order to detect variation within the core genome (genes shared by all organisms) or the accessory genome (genes present in only some organisms) [60].

The first description of *C. difficile* PCR RT027 phylogeny using high-throughput WGS demonstrated that 25 PCR RT027 isolates from the US and Europe could be further discriminated into 25 distinct genotypes based on SNP analysis [54]. Furthermore, this study demonstrated that isolates from different regions of the US and Europe occupy distinct evolutionary lineages and harbour unique antibiotic resistance genes. More recently, it was demonstrated that PCR RT027 isolates emerged through two distinct epidemic lineages after acquiring the same antibiotic resistance mutation; moreover these two lineages displayed different patterns of global spread [61]. The routine use of WGS in diagnostics and epidemiology is nicely reflected by the study of Koser et al. [62]. In this study it was reported that whole-genome SNP typing can be mainly used for monitoring outbreaks and recognition of pathogen transmission pathways. Current methods for monitoring *C. difficile* hospital associated outbreaks, such as PCR ribotyping, have too limited discriminatory power to characterise potential outbreak strains as the same bacterial clone. Sequencing of whole genomes offers the optimal discriminatory power allowing laboratories to detect transmission pathways between hospitals, hospital wards and patients on the same ward.

In addition, Eyre et al. [19] demonstrated that WGS can produce practical, clinically relevant data in a time frame that can influence patient management and infection control practice during an outbreak. Moreover, this study demonstrated that a cluster of healthcare-associated *C. difficile* cases caused by the same ST was in fact a number of unrelated sub-lineages, therefore allowing to rule out in patient-to-patient transmission. Furthermore, WGS combined with comparative genomics is an effective approach to identify novel genetic markers that are potentially linked to virulence. This is an important advantage above conventional typing methods that use existing markers for characterisation of isolates. Whole genome sequencing is not likely to replace routine diagnostic techniques in reference laboratories. For example, matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (TOF), which is rapid and easy to perform, is currently used in the Dutch reference laboratory for primary detection of pathogens.

In order to determine whether sequenced isolates are part of an outbreak, it must be defined how many SNP differences still represent 'related' isolates. For that reason, we should be informed on the rate of SNP accumulation in *C. difficile* lifecycle (molecular clock), although bacterial isolates with a hypermutator phenotype could complicate the determination of such a threshold [56]. The molecular clock rate of *C. difficile* was reported at 2.3 SNPs/genome/year in the study done by Eyre et al. [19]. Further study is necessary to confirm this rate of *C. difficile* evolution.

Application of typing methods to study the epidemiology of *Clostridium difficile* infections

An obvious reason to type *C. difficile* isolates is to early detect and investigate outbreaks, which can be defined as 'a temporal increase in the incidence of a bacterial species caused by transmission of a certain strain' [63]. In addition, typing methods contribute to epidemiological surveillance on national, European or worldwide level and can be used to report the incidence of various *C. difficile* types and recognise newly emerging virulent types [63]. Typing might also establish the local and global spread of bacteria and elucidate routes of transmission.

In the beginning of the 21st century, a worldwide increase in the incidence of CDI was seen. Soon thereafter, it was recognised that a specific type of *C. difficile*, PCR RT027, was linked to this increase of incidence [7,9]. PCR RT027 was associated with specific predisposing factors, course and outcome of CDI. In a large Canadian outbreak, fluoroquinolones were associated with PCR RT027 and mortality rates among patients with this type increased to 23% within 30 days of diagnosis [9,64]. In the Netherlands, molecular typing of *C. difficile* using PCR ribotyping contributed to recognition of an outbreak of two simultaneously occurring PCR RTs (027 and 017) [45]. Again, patients had PCR RT-specific risk factors and mortality rates. Numerous studies demonstrated the increased virulence of PCR RT027 [6–10] and found that other emerging types, such as PCR RT078, were also associated with specific risk factors or complicated clinical course [11]. Without results from typing methods, these associations would have stayed unrecognised.

Molecular typing results can also be used to compare the distribution of various *C. difficile* types isolated from animals, humans and food, which can hint towards food-borne disease or zoonotic potential of specific PCR RTs. The emerging *C. difficile* PCR RT078 in humans is found in high numbers in animals, especially piglets and calves [11,65–67]. Koene et al. [68] investigated the presence and characteristics of *C. difficile* in seven different animal species. PCR RTs 012, 014 and 078 were most frequently isolated among these Dutch animals, similar types were found among hospitalised patients in the Netherlands in 2009/2010. Meat consumption has also been suspected to contribute to transmission of *C. difficile*. PCR RTs 001, 017, 012 and 087 have been isolated from meat in Europe, however, isolation rates are low and might not be high enough to exceed the infectious dose [65–69] . Although PCR RTs in animals, meat and humans overlap, PCR ribotyping lacks discriminatory power to show clonal spread of *C. difficile* isolates from humans to animals. New

molecular methods should be developed and applied. The optimised MLVA scheme developed by Bakker et al. [44] showed relatedness between human and porcine PCR RT078 strains, although this could not always be confirmed with epidemiological data. Hopefully, highly discriminative typing methods such as whole-genome SNP typing can provide us with novel insights on zoonotic transmission.

Importance of molecular typing for national surveillance by reference laboratories

In Europe and North America, surveillance studies to monitor the incidence of CDI and the spread of hypervirulent strains have been established at regional and national levels since 2007 although reporting of CDI is not mandatory in all European Union (EU) countries. To enhance surveillance for CDI, the ECDC and the US CDC advised to widely launch surveillance programmes for CDI [28]. Consequently, a European network to support capacity building for standardised surveillance of CDI was initiated by the ECDC [28].

When methods and data on existing national CDI surveillance systems in Europe were reviewed (personal communication, A Kola, 2012), surveillance of CDI was reported in 45% (14/31) of the European countries. Active surveillance of CDI is performed in Austria, Norway, Belgium, Denmark, France, Germany, Ireland, Hungary, the Netherlands, Spain, Sweden, Luxembourg and the UK [46,70–79]. Surveillance was mostly continuous and prospective, but only four surveillance systems combined microbiological and epidemiological data (typing and susceptibility testing results) on a regular basis. A second recently completed survey in Europe (personal communication, D W Notermans, 2012) demonstrated that the majority of the laboratories were able to culture, but only half had access to typing. This limited typing capacity demonstrates the uncertainty of the true incidence levels of *C. difficile* types across Europe and hampers recognition of new emerging *C. difficile* types.

The contribution of national reference laboratories to survey CDI on a national level is illustrated by examples from the Netherlands and the UK. In 2005, soon after the emergence of *C. difficile* PCR RT027, the Center for Infectious Disease Control (CIb) of the National Institute for Public Health and the Environment (RIVM) in the Netherlands started a national Reference Laboratory for *C. difficile*. In 2009, this laboratory noticed an emergence of a new virulent PCR RT078, which was the third most frequently found type in the Netherlands among humans and was present in nearly all pig farms investigated [11,67]. Subsequently, this type was also found emerging in other European countries [80]. Recently, the reference laboratory noticed a re-emergence of *C. difficile* PCR RT027 since 2010. In the period between May 2011 and May 2012, 289 samples from 26 healthcare facilities and laboratories in the Netherlands were submitted because of severe CDI cases or outbreaks. PCR RTs 001 and 027 were the most commonly found (both 15.0%). Interestingly, in contrast to a previous report of declining PCR RT027 in hospitals in the Netherlands [81], type 027 was frequently identified in long-term care facilities associated with exchange of patients to neighbouring hospitals.

In the UK, the *C. difficile* Ribotyping Network (CDRN) was established in 2007, as part of improved CDI surveillance, to facilitate the detection and control of epidemic strains. Between 2007 and 2010, the CDRN received a large number of isolates (n=11,294) for PCR ribotyping. Typing results indicated that almost all of the 10 most common PCR RTs changed significantly during this time period [79]. As the proportion of CDI caused by PCR RT027 declined (from 55% to 21%), significant increases were observed in the prevalence of other *C. difficile* types, especially PCR RTs 014/020, 015, 002, 078, 005, 023, and 016. In addition, there was a 61% reduction in reports of *C. difficile* in England from 2008 to 2011, which occurred coincidently as the proportion of CDI caused by *C. difficile* PCR RT027 declined. Notably, the large reduction in incidence of *C. difficile* PCR RT027 cases has been paralleled by decreases in CDI related mortality [82]. The perceived success of the surveillance programme means that currently approximately a third of all CDI cases in England are referred to CDRN. CDI control programs should ideally include prospective access to *C. difficile* typing and analysis of risk factors for CDI and outcomes.

Future perspective

In the last fifteen years molecular genotyping methods have replaced some of the more traditional typing methods. WGS will dominate the field of molecular typing in the next decade. However, before WGS can be used as a routine tool for molecular typing some requirements need to be fulfilled. First, WGS needs to be fast, preferentially within 48 hours. Furthermore, the technical workflow including data analysis needs to be simplified into an automatic pipeline. Finally, the costs for acquiring the technical and organisational platform needed to perform WGS must be reduced. Fulfilling, these requirements, which is in our opinion a matter of time, would greatly increase the use of WGS worldwide.

Acknowledgments

This work was supported by ZonMw Grant 50-50800-98-079 from the Netherlands Organization for Scientific Research (NWO).

We would like to acknowledge the European Study group of *Clostridium difficile* on behalf of the European Society for Clinical Microbiology and Infectious Diseases (E.J.K.) for their contribution.

This work was funded by the European Centre for Disease Prevention and Control (ECDC) through the call for tender OJ/2010/07/09-PROC/2010/035.

References

1. Bartlett JG. Antibiotic-associated pseudomembranous colitis. Hosp Pract (Off Ed). 1981;16(12):85-8, 93-5.
2. Kelly CP, Pothoulakis C, LaMont JT. Clostridium difficile colitis. N Engl J Med. 1994;330(4):257-62.
3. Voth DE, Ballard JD. Clostridium difficile toxins: mechanism of action and role in disease. Clin Microbiol Rev. 2005;18(2):247-63.
4. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of toxin A and toxin B in Clostridium difficile infection. Nature. 2010;467 (7316):711-3.
5. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, et al. Toxin B is essential for virulence of Clostridium difficile. Nature. 2009;458(7242):1176-9.
6. Schwan C, Stecher B, Tzivelekidis T, van Ham M, Rohde M, Hardt WD, et al. Clostridium difficile toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. PLoS Pathog. 2009;5(10):e1000626.
7. Kuijper EJ, Coignard B, Tüll P; ESCMID Study Group for Clostridium difficile; EU Member States; European Centre for Disease Prevention and Control. Emergence of Clostridium difficile-associated disease in North America and Europe. Clin Microbiol Infect. 2006;12 Suppl 6:2-18.
8. McDonald LC, Killgore GE, Thompson A, Owens RC Jr, Kazakova SV, Sambol SP, et al. An epidemic, toxin gene-variant strain of Clostridium difficile. N Engl J Med. 2005;353(23):2433-41.
9. Pépin J, Valiquette L, Cossette B. Mortality attributable to nosocomial Clostridium difficile-associated disease during an epidemic caused by a hypervirulent strain in Quebec. CMAJ. 2005;173(9):1037-42.
10. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, et al. Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. Lancet. 2005;366(9491):1079-84.
11. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, et al. Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis. 2008;47(9):1162-70.
12. Cohen SH, Tang YJ, Silva J Jr. Molecular typing methods for the epidemiological identification of Clostridium difficile strains. Expert Rev Mol Diagn. 2001;1(1):61-70.
13. Delmee M, Homel M, Wauters G. Serogrouping of Clostridium difficile strains by slide agglutination. J Clin Microbiol. 1985;21(3):323-7.
14. Tabaqchali S, Holland D, O'Farrell S, Silman R. Typing scheme for Clostridium difficile: its application in clinical and epidemiological studies. Lancet. 1984;1(8383):935-8.
15. Toma S, Lesiak G, Magus M, Lo HL, Delmée M. Serotyping of Clostridium difficile. J Clin Microbiol. 1988;26(3):426-8.
16. Mulligan ME, Peterson LR, Kwok RY, Clabots CR, Gerding DN. Immunoblots and plasmid fingerprints compared with serotyping and polyacrylamide gel electrophoresis for typing Clostridium difficile. J Clin Microbiol. 1988;26(1):41-6.
17. Kuijper EJ, van den Berg RJ, Brazier JS. Comparison of molecular typing methods applied to Clostridium difficile. Methods Mol Biol. 2009;551:159-71.
18. Killgore G, Thompson A, Johnson S, Brazier J, Kuijper E, Pepin J, et al. Comparison of seven techniques for typing international epidemic strains of Clostridium difficile: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. J Clin Microbiol. 2008;46 (2):431-7.
19. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open. 2012; 2(3).
20. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
21. Gürtler V. Typing of Clostridium difficile strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. J Gen Microbiol. 1993;139(12):3089-97.
22. Bidet P, Barbut F, Lalande V, Burghoffer B, Petit JC. Development of a new PCR-ribotyping method for Clostridium difficile based on ribosomal RNA gene sequencing. FEMS Microbiol Lett. 1999;175(2):261-6.
23. Stubbs SL, Brazier JS, O'Neill GL, Duerden BI. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of Clostridium difficile and construction of a library consisting of 116 different PCR ribotypes. J Clin Microbiol. 1999;37(2):461-3.
24. van den Berg RJ, Claas EC, Oyib DH, Klaassen CH, Dijkshoorn L, Brazier JS, et al. Characterization of toxin A-negative, toxin B-positive Clostridium difficile isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping. J Clin Microbiol. 2004; 42(3):1035-41.
25. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988;26(11):2465-6.
26. Kristjánsson M, Samore MH, Gerding DN, DeGirolami PC, Bettin KM, Karchmer AW, et al. Comparison of restriction endonuclease analysis, ribotyping, and pulsed-field gel electrophoresis for molecular differentiation of Clostridium difficile strains. J Clin Microbiol. 1994;32(8):1963-9.
27. Centers for Disease Control and Prevention (CDC). PulseNet. Atlanta: CDC. [Accessed 1 Jul 2012]. Available from: http://www.cdc.gov/pulsenet/
28. European Clostridium difficile infection study network (ECDIS-NET). Europe: Supporting capacity building for surveillance of Clostridium difficile. Leiden: ECDIS-NET. [Accessed 1 Jul 2012]. Available from: http://www.ecdisnet.eu/
29. Brazier JS, Delmee M, Tabaqchali S, Hill LR, Mulligan ME, Riley TV. Proposed unified nomenclature for Clostridium difficile typing. Lancet. 1994;343(8912):1578-9.
30. Lemee L, Dhalluin A, Pestel-Caron M, Lemeland JF, Pons JL. Multilocus sequence typing analysis of human and animal Clostridium difficile isolates of various toxigenic types. J Clin Microbiol. 2004;42(6):2609-17.
31. Clostridium difficile MLST Databases. Clostridium difficile Multi Locus Sequence Typing website. Oxford: University of Oxford. [Accessed 21 nov 2012]. Available from: http://pubmlst.org/cdifficile/
32. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, et al. Multilocus sequence typing of Clostridium difficile. J Clin Microbiol. 2010;48(3):770-8.
33. Institut Pasteur. Clostridium difficile MLST Database. Paris: Institut Pasteur. [Accessed: 1 jul 2012]. Available from: http://www.pasteur.fr/recherche/genopole/PF8/mlst/Cdifficile2.html
34. Indra A, Blaschitz M, Kernbichler S, Reischl U, Wewalka G, Allerberger F. Mechanisms behind variation in the Clostridium difficile 16S-23S rRNA intergenic spacer region. J Med Microbiol. 2010;59(Pt 11):1317-23.
35. Dingle KE, Griffiths D, Didelot X, Evans J, Vaughan A, Kachrimanidou M et al. Clinical Clostridium difficile: clonality and pathogenicity locus diversity. PLoS One. 2011; 6(5):e19993.
36. Stabler RA, Dawson LF, Valiente E, Cairns MD, Martin MJ, Donahue EH, et al. Macro and micro diversity of Clostridium difficile isolates from diverse sources and geographical locations. PLoS One. 2012;7(3):e31559.
37. Knetsch CW, Terveer EM, Lauber C, Gorbalenya AE, Harmanus C, Kuijper EJ, et al. Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. Infect Genet Evol. 2012;12(7):1577-85.
38. Eckert C, Vromman F, Halkovich A, Barbut F. Multilocus variable-number tandem repeat analysis: a helpful tool for subtyping French Clostridium difficile PCR ribotype 027 isolates. J Med Microbiol. 2011; 60(Pt 8):1088-94.
39. Fawley WN, Freeman J, Smith C, Harmanus C, van den Berg RJ, Kuijper EJ, et al. Use of highly discriminatory fingerprinting to analyze clusters of Clostridium difficile infection cases due to epidemic ribotype 027 strains. J Clin Microbiol. 2008;46(3):954-60.
40. Lindstedt BA. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis. 2005;26(13):2567-82.
41. Marsh JW, O'Leary MM, Shutt KA, Pasculle AW, Johnson S, Gerding DN, et al. Multilocus variable-number tandem-repeat analysis for investigation of Clostridium difficile transmission in Hospitals. J Clin Microbiol. 2006;44(7):2558-66.
42. van den Berg RJ, Schaap I, Templeton KE, Klaassen CH, Kuijper EJ. Typing and subtyping of Clostridium difficile isolates by using multiple-locus variable-number tandem-repeat analysis. J Clin Microbiol. 2007;45(3):1024-8.
43. Broukhanski G, Low DE, Pillai DR. Modified multiple-locus variable-number tandem-repeat analysis for rapid identification and typing of Clostridium difficile during institutional outbreaks. J Clin Microbiol. 2011;49(5):1983-6.
44. Bakker D, Corver J, Harmanus C, Goorhuis A, Keessen EC, Fawley WN, et al. Relatedness of human and animal Clostridium difficile PCR Ribotype 078 isolates based on multilocus variable-number tandem-repeat analysis and tetracycline resistance. J Clin Microbiol. 2010;48(10):3744-9.
45. Goorhuis A, Debast SB, Dutilh JC, van Kinschot CM, Harmanus C, Cannegieter SC, et al. Type-specific risk factors and outcome

in an outbreak with 2 different Clostridium difficile types simultaneously in 1 hospital. Clin Infect Dis. 2011;53(9):860-9.

46. Fawley WN, Wilcox MH; Clostridium difficile Ribotyping Network for England and Northern Ireland. An enhanced DNA fingerprinting service to investigate potential Clostridium difficile infection case clusters sharing the same PCR ribotype. J Clin Microbiol. 2011;49(12):4333-7.
47. Janvilisri T, Scaria J, Thompson AD, Nicholson A, Limbago BM, Arroyo LG, et al. Microarray identification of Clostridium difficile core components and divergent regions associated with host origin. J Bacteriol. 2009;191(12):3881-91.
48. Manzoor SE, Tanner HE, Marriott CL, Brazier JS, Hardy KJ, Platt S, et al. Extended multilocus variable-number tandem-repeat analysis of Clostridium difficile correlates exactly with ribotyping and enables identification of hospital transmission. J Clin Microbiol. 2011;49(10):3523-30.
49. Wei HL, Kao CW, Wei SH, Tzen JT, Chiou CS. Comparison of PCR ribotyping and multilocus variable-number tandem-repeat analysis (MLVA) for improved detection of Clostridium difficile. BMC Microbiol. 2011;11:217.
50. Indra A, Huhulescu S, Schneeweis M, Hasenberger P, Kernbichler S, Fiedler A, et al. Characterization of Clostridium difficile isolates using capillary gel electrophoresis-based PCR ribotyping. J Med Microbiol. 2008;57(Pt 11):1377-82.
51. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331(6016):430-4.
52. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010;327(5964):469-74.
53. Harris SR, Clarke IN, Seth-Smith HM, Solomon AW, Cutcliffe LT, Marsh P, et al. Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet. 2012;44(4):413-9.
54. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci U S A. 2010;107(16):7527-32.
55. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet. 2008;40(8):987-93.
56. Köser 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.
57. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
58. Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, et al. Genomics. Genome project standards in a new era of sequencing. Science. 2009;326(5950):236-7.
59. Wellcome Trust Sanger Institute. Clostridium difficile genome data. Hinxton: Wellcome Trust Sanger Institute. [Accessed: 1 Nov 2012]. Available from: http://www.sanger.ac.uk/resources/downloads/bacteria/clostridium-difficile.html
60. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Curr Opin Genet Dev. 2005;15(6):589-94.
61. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet. 2012;45(1):109-13.
62. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 2012;8(8):e1002824.
63. 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.
64. Pépin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, et al. Emergence of fluoroquinolones as the predominant risk factor for Clostridium difficile-associated diarrhea: a cohort study during an epidemic in Quebec. Clin Infect Dis. 2005;41(9):1254-60.
65. Hensgens MP, Keessen EC, Squire MM, Riley TV, Koene MG, de Boer E, et al. Clostridium difficile infection in the community: a zoonotic disease? Clin Microbiol Infect. 2012;18(7):635-45.
66. Jhung MA, Thompson AD, Killgore GE, Zukowski WE, Songer G, Warny M, et al. Toxinotype V Clostridium difficile in humans and food animals. Emerg Infect Dis. 2008;14(7):1039-45.
67. Keessen EC, Gaastra W, Lipman LJ. Clostridium difficile infection in humans and animals, differences and similarities. Vet Microbiol. 2011;153(3-4):205-17.
68. Koene MG, Mevius D, Wagenaar JA, Harmanus C, Hensgens MP, Meetsma AM, et al. Clostridium difficile in Dutch animals: their presence, characteristics and similarities with human isolates. Clin Microbiol Infect. 2012;18(8):778-84.
69. de Boer E, Zwartkruis-Nahuis A, Heuvelink AE, Harmanus C, Kuijper EJ. Prevalence of Clostridium difficile in retailed meat in the Netherlands. Int J Food Microbiol. 2011. 144(3):561-4.
70. Asensio A, Vaque-Rafart J, Calbo-Torrecillas F, Gestal-Otero JJ, Lopez-Fernandez F, Trilla-Garcia A, et al. Increasing rates in Clostridium difficile infection (CDI) among hospitalised patients, Spain 1999-2007. Euro Surveill. 2008;13(31):pii=18943. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18943
71. Birgand G, Blanckaert K, Carbonne A, Coignard B, Barbut F, Eckert C, et al. Investigation of a large outbreak of Clostridium difficile PCR-ribotype 027 infections in northern France, 2006-2007 and associated clusters in 2008-2009. Euro Surveill. 2010;15(25):pii=19597. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19597
72. Indra A, Huhulescu S, Fiedler A, Kernbichler S, Blaschitz M, Allerberger F. Outbreak of Clostridium difficile 027 infection in Vienna, Austria 2008-2009. Euro Surveill. 2009;14(17):pii=19186. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19186
73. Ingebretsen A, Hansen G, Harmanus C, Kuijper EJ. First confirmed cases of Clostridium difficile PCR ribotype 027 in Norway. Euro Surveill. 2008;13(2):pii=8011. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=8011
74. Kleinkauf N, Weiss B, Jansen A, Eckmanns T, Bornhofen B, Küehnen E, et al. Confirmed cases and report of clusters of severe infections due to Clostridium difficile PCR ribotype 027 in Germany. Euro Surveill. 2007;12(46):pii=3307. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=3307
75. Kotila SM, Virolainen A, Snellman M, Ibrahem S, Jalava J, Lyytikäinen O. Incidence, case fatality and genotypes causing Clostridium difficile infections, Finland, 2008. Clin Microbiol Infect. 2011;17(6):888-93.
76. Suetens C. Clostridium difficile: summary of actions in the European Union. Euro Surveill. 2008;13(31):pii=18944. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18944
77. Terhes G, Urbán E, Konkoly-Thege M, Székely E, Brazier JS, Kuijper EJ, et al. First isolation of Clostridium difficile PCR ribotype 027 from a patient with severe persistent diarrhoea in Hungary. Clin Microbiol Infect. 2009;15(9):885-6.
78. Viseur N, Lambert M, Delmee M, Van Broeck J, Catry B (2011) Nosocomial and non-nosocomial Clostridium difficile infections in hospitalised patients in Belgium - compulsory surveillance data from 2008 to 2010. Euro Surveill. 16(43):pii=20000. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20000
79. Wilcox MH, Shetty N, Fawley WN, Shemko M, Coen P, Birtles A, et al. Changing epidemiology of Clostridium difficile infection following the introduction of a national ribotyping based surveillance scheme in England. Clin Infect Dis. 2012;55(8):1056-63.
80. Bauer MP, Notermans DW, van Benthem BH, Brazier JS, Wilcox MH, Rupnik M, et al. Clostridium difficile infection in Europe: a hospital-based survey. Lancet. 2011;377(9759):63-73.
81. Hensgens MP, Goorhuis A, Notermans DW, van Benthem BH, Kuijper EJ. Decrease of hypervirulent Clostridium difficile PCR ribotype 027 in the Netherlands. Euro Surveill. 2009;14(45):pii=19402. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19402
82. Office for National Statistics (ONS). Deaths Involving Clostridium Difficile - England and Wales, 2006 to 2010. Newport: ONS. [Accessed: 1 Jul 2012]. Available from: http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-clostridium-difficile/2006-to-2010/statistical-bulletin.html

Within-patient emergence of the influenza A(H1N1) pdm09 HA1 222G variant and clear association with severe disease, Norway

R Rykkvin[1], A Kilander (anette.kilander@fhi.no)[1], S G Dudman[1], O Hungnes[1]
1. Department of Virology, Norwegian Institute of Public Health, Oslo, Norway

Citation style for this article:
Rykkvin R, Kilander A, Dudman SG, Hungnes O. Within-patient emergence of the influenza A(H1N1)pdm09 HA1 222G variant and clear association with severe disease, Norway . Euro Surveill. 2013;18(3):pii=20369. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20369

Article submitted on 29 June 2012 / published on 17 January 2013

The association between a particular mutation in the HA1 subunit of the influenza virus haemagglutinin, D222G, and severe and fatal disease in cases of influenza A(H1N1)pdm09 in Norway during the 2009 pandemic was investigated using pyrosequencing. The prevalence of the variant among fatal cases was 8/26 and among severe non-fatal cases 5/52. No D222G mutations were found among the 381 mild cases. This difference could not be attributed to sampling differences, such as body location of sampling, or duration of illness. In cases with mutant virus where clinical specimens from different days of illness were available, transition from wild-type to mutant virus was commonly observed (4/5), indicating that the mutant virus emerged sporadically in individual patients. In patients with paired samples from both the upper and lower respiratory tract (n=8), the same viral genotypes were detected in both locations. In most of the D222G cases (11/13), the mutant virus was found as a quasispecies.

Introduction

Infection with the pandemic influenza A(H1N1) virus that emerged in 2009 led to mild disease in the vast majority of cases; however, there was also an unusual occurrence of viral pneumonia, severe disease and death in younger age groups than commonly observed for seasonal influenza [1]. In a large proportion of severe cases, conditions predisposing for severe disease have been identified [2] and host factors therefore appear to strongly influence the clinical outcome of infection. On the other hand, this novel virus of zoonotic origin differed from the previous seasonal A(H1N1) virus in the resulting disease profile; thus, viral determinants of pathogenicity must also be involved – e.g. it has been shown to be more pneumotropic than seasonal A(H1N1) virus in a ferret model [3]. It is important to understand better what viral and host-related factors determine the observed dichotomous pathogenicity profile.

The first cases of influenza A(H1N1)pdm09 virus infection in Norway were recorded in early May 2009, shortly after emergence in Mexico, but cases were few and scattered across the country until mid-summer. A limited influenza epidemic took place in late July/early August, followed by a comparatively calm period leading up to a major epidemic that surpassed all previous peaks recorded in the current surveillance system. The epidemic reached its highest point in early November and by the end of 2009, it had mostly subsided.

As part of the intensified surveillance carried out during the 2009 influenza pandemic, the national reference laboratory for human influenza at the Norwegian Institute of Public Health received a large number of respiratory specimens from confirmed and possible cases of influenza A(H1N1)pdm09. In late November 2009, we noticed that a particular amino acid substitution – aspartic acid (D) to glycine (G) in the viral haemagglutinin (HA) glycoprotein subunit HA1 at position 222 (D222G) – appeared in fatal cases, while we did not find it in the numerous mild cases analysed. Realising that a similar pattern seemed to be taking place in the Ukraine (R. Daniels, personal communication, November 2009) and given that the mutation had been shown to influence viral receptor specificity in another influenza A(H1N1) virus [4], it was decided to notify international public health authorities and other national authorities about this possible pathogenicity determinant, to expedite assessment of it [5]. A preliminary review in January 2010 of D222G amino acid substitution in the HA of influenza A(H1N1)pdm09 viruses from the World Health Organization (WHO) stated that mutations, including those leading to the D222G substitution in the HA, had appeared sporadically since the first emergence of influenza A(H1N1)pdm09 viruses, and that the substitutions in HA had been reported in viruses obtained from cases of mild to severe to fatal illnesses but that such viruses had neither formed distinct phylogenetic clustering nor been associated with consistent changes in virus antigenicity [6].

Since the first account of our findings in Norway [7], investigations into the occurrence of this mutation resulted in an increasing number of reports [8-28]. While the prevalence of this mutation varied between

the reporting countries, in most studies, the 222G mutation is primarily found in severe and fatal cases. One of the first larger studies came from a group in Hong Kong, who analysed this amino acid position in severe and non-severe cases of influenza A(H1N1) pdm09 [13]. Nine (4.1%) of 219 severe or fatal cases of pandemic influenza had the mutation, in contrast to none of 239 non-severe cases.

Data from these reports indicated that the D222G mutation was absent or uncommon in viruses that were in sustained circulation. However, one case of transmission of a 222G virus was reported [9], but the transmitted virus in this case had acquired an additional mutation that may have influenced receptor binding characteristics.

To further investigate the sporadic occurrence of the 222G mutant influenza virus, we performed a more in-depth analysis of an expanded data set. Our original data set [7] included 266 cases, while the expanded set comprised 462 cases. The present study included assessment of the majority of the fatal cases in Norway, as well as of a larger number of samples from severe non-fatal and mild cases collected throughout the pandemic.

We studied the prevalence of HA1 222 mutations within different clinical outcome groups, in serially collected specimens, in upper versus lower respiratory tract and in early versus late specimens. We also analysed age and sex distribution and examined the characteristics of the fatal cases.

We further compared the mutant viruses phylogenetically, looked for the presence of mutant quasispecies and oseltamivir resistance.

Methods

Data and clinical materials

As part of the intensified surveillance instigated in response to the emergence of influenza A(H1N1)pdm09 in April 2009, virus-containing specimens from all parts of the country were received in the Norwegian Institute of Public Health, which serves as the National Influenza Centre in Norway. A total of 15 medical microbiology laboratories submitted specimens containing influenza virus (original specimens, nucleic acid preparations from original specimens or virus isolates) to the National Influenza Centre for further characterisation. Most of these patient specimens originated from primary care clinics; the remainder were from hospitals.

Each laboratory sent a maximum of five specimens each week plus any specimens from patients with severe disease, suspicion of antiviral resistance or other special circumstances such as suspected vaccine failure. In addition, intensive care and fatal cases remained notifiable to the Norwegian Institute of Public Health throughout the pandemic and we actively solicited materials from microbiology and pathology laboratories that were in possession of specimens from these cases. Together, these collection schemes enabled us to obtain specimens from nearly all the recorded fatal influenza A(H1N1)pdm09 cases in Norway as well as a large number of severe and mild cases. Patient information relevant for this study was obtained primarily from the patient referral forms that came with the specimens, supplemented with information gathered in the notification of severe and fatal cases and, in a few cases, from direct contact with clinicians who cared for the patient. Cases were assigned to clinical outcome groups by a medical specialist, according to WHO guidance criteria [2]. Briefly, the criteria for complicated/severe influenza included: clinical and/or radiological signs of lower respiratory tract disease, central nervous system involvement, severe dehydration, secondary complications (renal failure, multiorgan failure, septic shock, rhabdomyolysis, myocarditis), exacerbation of underlying chronic disease or signs and symptoms of progressive influenza disease. The criteria were modified as follows: hospital admission in itself did not lead to classification as severe influenza, without additional evidence of complicated disease. This modification was necessary due to a low threshold for hospitalisation of patients during parts of the pandemic. Furthermore, sustained virus replication in itself was not regarded as evidence of severe influenza. From the available specimens, we picked all the specimens from fatal and ICU cases as well as a large subset of specimens from the other cases, making sure that viruses from all parts of Norway and from the entire first period of A(H1N1)pdm09 virus circulation in the country (May 2009–January 2010) were well represented. The selected specimens were sequenced with regard to the codon encoding amino acid 222 of the HA1 subunit. Only cases where the HA1 222 genotype could be ascertained in the original specimen were included in the study.

Detection of HA1 222 mutations at nucleic acid level

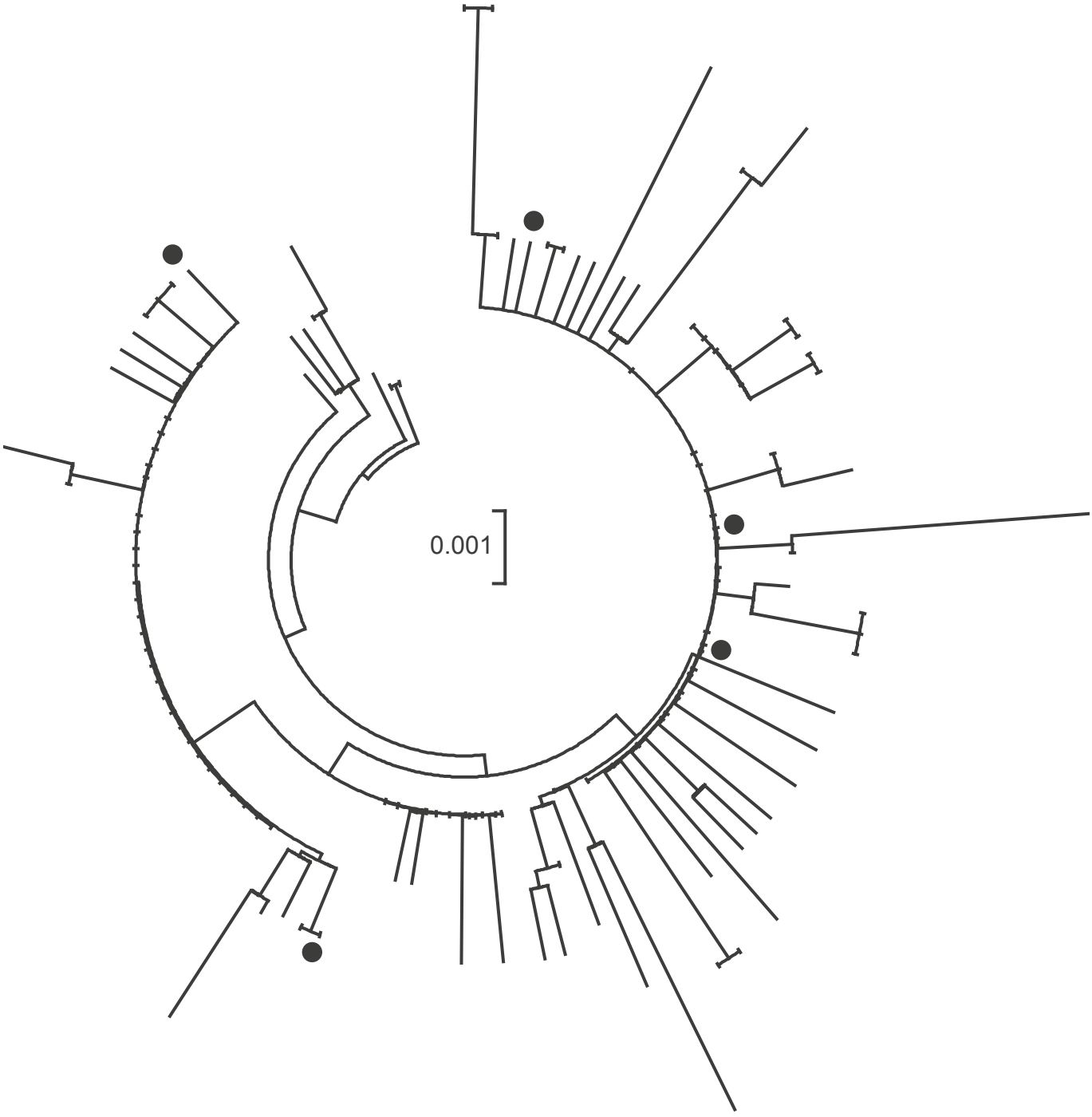
Initially, mutations were detected through conventional (Sanger) cycle sequencing of reverse transcription-polymerase chain reaction (RT-PCR) amplicons obtained from viral RNA from virus isolates and primary specimens.

Viral RNA was extracted using a total nucleic acid extraction kit in the MagNA Pure LC System (Roche Diagnostics, Mannheim, Germany). In general, a modification of the full genome sequencing protocol provided by the United States Centers for Disease Control and Prevention [28] was used for virus isolates, whereas sequencing of the amplicon from a more sensitive RT-PCR [30] was used for many of the primary specimens.

As soon as we became aware of the possible importance of the mutation, a pyrosequencing assay was designed and used. Briefly, a 110-nucleotide amplicon

Figure 1

Phylogenetic reconstruction of influenza A(H1N1)pdm09 haemagglutinin gene (HA1), Norway 2009/10, showing the distribution of 222G viruses across the phylogeny



Viruses harbouring 222G are marked with solid circles. The entire coding region of the HA1 subunit was analysed. The codon for amino acid 222 was excluded for this analysis, in order to investigate how 222G mutants were otherwise related.

Table 1

Risk factors for severe illness in fatal cases of influenza A(H1N1)pdm09 with available haemagglutinin gene (HA1) 222 genotype, Norway, 2009/10 (n=26)

Risk factors for severe disease	Number of cases
Presence of risk factors	
None	4
One or more	21
Data unavailable	1
Individual risk factors[a]	
Chronic respiratory disease	4
Chronic heart disease	7
Chronic renal or hepatic disease	0
Diabetes	2
Pregnancy	1
Obesity[b]	3
Immunosuppression	3
Chronic neurological disease or injury	3
Other	7

[a] Each patient may have several risk factors.

[b] Obesity: for one of these patients, a body mass index (BMI) of >40 was reported. The other two were reported as 'obese' or 'adipose', with no BMI given.

encompassing the region of the HA gene which encodes amino acid 222 was generated from 5 µl specimen RNA combined with each of primers pyro-H1 forward: 5'-AGTTCAAGCCGGAAATAGCA-3' and pyro-H1 reverse: 5'-biotin-TTTCCAGTTGCTTCGAATGTT-3' and reagents from the SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity kit (Invitrogen) in a 25 µl reaction and subjected to the following cycling conditions: 30 min. at 50 °C, 2 min. at 94 °C; 45 cycles of 15 sec. denaturation at 94 °C, 30 sec. annealing at 55 °C and 1 min. extension at 68 °C; finally, 5 min. at 68 °C for final extension.

The pyrosequencing reactions were performed as previously described [29] using a residue-specific sequencing primer (5'-AGCAATAAGACCCAAAGTGAGG-3').

The sequenced region begins at nucleotide 747 in the full-length sequence of the viral RNA segment 4, where nucleotides 747–749 encode HA1 amino acid 222. The most common wild-type codon in this position is GAT, which encodes aspartic acid (D). The GAA codon encodes glutamic acid (E), while GGT (or GGA) gives glycine (G). Mutation of G to A in the first position of the GAT codon gives asparagine (N).

The viruses were also analysed for oseltamivir susceptibility by detecting the H274Y mutation by sequence analysis, through either a pyrosequencing assay targeting the relevant single-point mutation [31] or through full- or partial-length cycle sequencing of the coding region for the viral neuraminidase.

Phylogenetic analysis

Phylogenetic analysis was performed on aligned sequences comprising the entire coding region of the HA1 subunit. Kimura 2-parameter corrected pairwise distances between manually aligned sequences were computed using the PHYLIP [32] program DNADIST, and the phylogenetic tree inferred using NEIGHBOR. The resulting tree was visualised using MEGA version 5.0 [33].

Statistical analysis

Statistical analysis was undertaken using PASW 17, version 17.02 (SPSS Inc., Chicago, United States). Descriptive statistics were calculated as medians with upper and lower range or as means with 95% confidence intervals. All hypothesis tests were two-tailed and statistical significance was assessed at the 0.05 level. Comparisons of categorical variables were performed using Fisher's exact test and comparisons of non-normal continuous distributions using the Mann–Whitney U test.

Results

In total, 462 influenza A(H1N1)pdm09 cases where the HA1 222 genotype could be ascertained in the primary specimen were included in the study. A large proportion of the isolated viruses were sent to the WHO Collaborative Centre for Reference and Research on influenza in the National Institute for Medical Research, Mill Hill, United Kingdom, for further characterisation. No changes in virus antigenicity were found (data not shown). We did not find the H274Y mutation, commonly associated with oseltamivir resistance, in any of the viruses in this study. In phylogenetic analysis of the HA1 coding region of Norwegian viruses (Figure 1), the five 222G harbouring viruses for which we could ascertain the full HA1 coding sequence occurred on different branches of the tree, showing that they did not form a distinct genetic cluster. This is in good agreement with the fact that the cases with 222G viruses occurred sporadically across Norway, at different times during the period of widespread virus circulation.

Of the 462 cases included in the analysis, 381 had mild disease, 52 had severe disease but were non-fatal and 26 were fatal (clinical outcome was unknown for three cases).

The median age of the cases with mild disease included was 20 years (range: 0–87), for the non-fatal cases with severe disease 27 years (range: 0–66 years) and for the fatal cases 38 years (range: 0–71). Among all 462 cases, 225 (49%) were men and 237 (51%) were women. The proportion of men versus women was similar for mild cases (48% vs 52%), severe non-fatal cases (52% vs 48%) and fatal cases (46% vs 54%). Of the 13 cases with 222G virus, eight were men and five were women: these proportions (62% vs 38%) were not

Table 2

Prevalence of haemagglutinin gene (HA1) 222 genotypes by clinical outcome (mild, severe non-fatal, fatal disease), influenza A(H1N1)pdm09 cases, Norway, 2009/10 (n=462)

HA1 222 genotype	Clinical outcome[a]					Total n (%)
	Mild n (%)	Severe non-fatal n (%)	Fatal n (%)	Severe including fatal n (%)	Unknown n (%)	
222D wild type	329 (86.4)	40 (76.9)	15 (57.7)	55 (70.5)	2 (66.7)	386 (83.5)
222G	0 (0)	5 (9.6)	8 (30.8)	13 (16.7)	0 (0)	13 (2.8)
222E	51 (13.4)	5 (9.6)	2 (7.7)	7 (9.0)	1 (33.3)	59 (12.8)
222N	1 (0.3)	2 (3.8)	1 (3.8)[b]	3 (3.8)[b]	0 (0)	4 (0.9)[b]
Total	**381 (100)**	**52 (100)**	**26 (100)**	**78 (100)**	**3 (100)**	**462 (100)**

[a] Severe non-fatal and fatal cases are shown separately and jointly.

[b] Additionally, one fatal 222G case harboured 222N virus as a quasispecies.

significantly different from those cases with non-222G viruses (48% vs 52%, n=449) (p=0.41, Fisher's exact test).

In the total data set, 13 cases harboured 222G viruses, 222E occurred in 59 cases and 222N in four cases (one additional case contained both 222G and 222N quasispecies, and was classified as the former). In 386 cases we detected only wild-type 222D viruses (Table 2). The presence of mutants found in the pyrosequencing assay was verified by conventional sequencing in all 10 cases where sequencing of primary samples was possible.

In order to avoid bias stemming from preferential testing of severe cases in our samples, the prevalence of different genetic variants was recorded within each clinical outcome group, instead of the other way round. Among the 26 fatal cases, eight harboured the 222G virus at some time during the course of infection. The corresponding proportion for severe non-fatal cases was 5/52 (10 %) while none was found in 381 analysed cases with mild disease. The proportions in fatal as well as in severe non-fatal cases were higher than in mild cases (p<0.0005). Furthermore, the higher frequency in fatal cases compared with severe non-fatal cases was also statistically significant (p=0.026, Fisher's exact test).

One specimen from a fatal case contained both 222G and 222N viruses in addition to wild-type 222D. Two of the 222G mutants had adenine in the third codon position and thus are likely to have arisen from 222E viruses. The prevalence of the 222E variant showed no significant difference between the various clinical outcome groups, and this variant represents a circulating clade with no apparent effect on virulence [9.10].

In the majority of 222G cases (11/13), the 222G mutants occurred as quasispecies, typically with wild type 222D sequence being more frequent (exemplified in Figure 2). The proportion of mutant virus in mixed populations as estimated by pyrosequencing tended to be less than 50% (range: 18–59). In all 10 cases analysed by both conventional sequencing and pyrosequencing, these mixtures were also evident as double peaks in conventional sequencing (see example in Figure 2).

In eight cases, paired specimens from upper and lower respiratory tract were available for analysis. Six of these cases had a fatal outcome. These specimen pairs were collected on the same day from the same patients. Upper/lower respiratory tract specimens from the same case but collected on different days were not considered as pairs in this analysis. Both 222D and 222G viruses were found in the samples, but in all eight pairs, the HA1 222 genotype in the upper and lower respiratory tract samples matched (Table 3).

Table 3

Prevalence of haemagglutinin gene (HA1) 222 genotypes in paired[a] upper and lower respiratory tract samples, influenza A(H1N1)pdm09 cases, Norway, 2009/10 (n=8)

HA1 222 genotype		Number of patients
Upper respiratory tract sample[b]	Lower respiratory tract sample[c]	
222D	222D	5
222G	222G	2
222D/G mix	222D/G mix	1

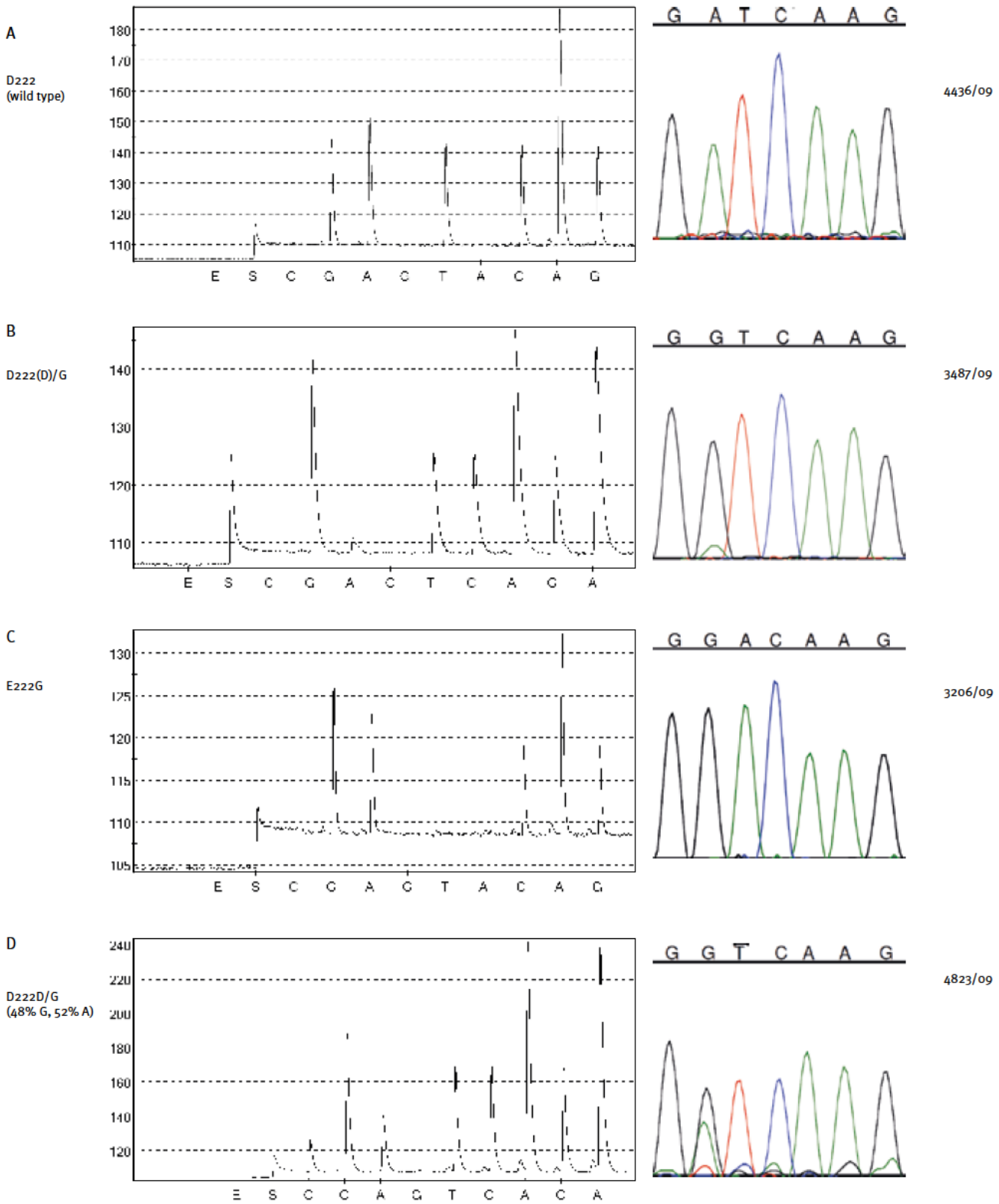
[a] The samples were collected from the same patients on the same day.

[b] Upper respiratory tract samples included nasopharyngeal swabs/aspirates, nasal swabs and throat swabs.

[c] Lower respiratory tract samples included tracheal aspirates and lung autopsy material.

Figure 2

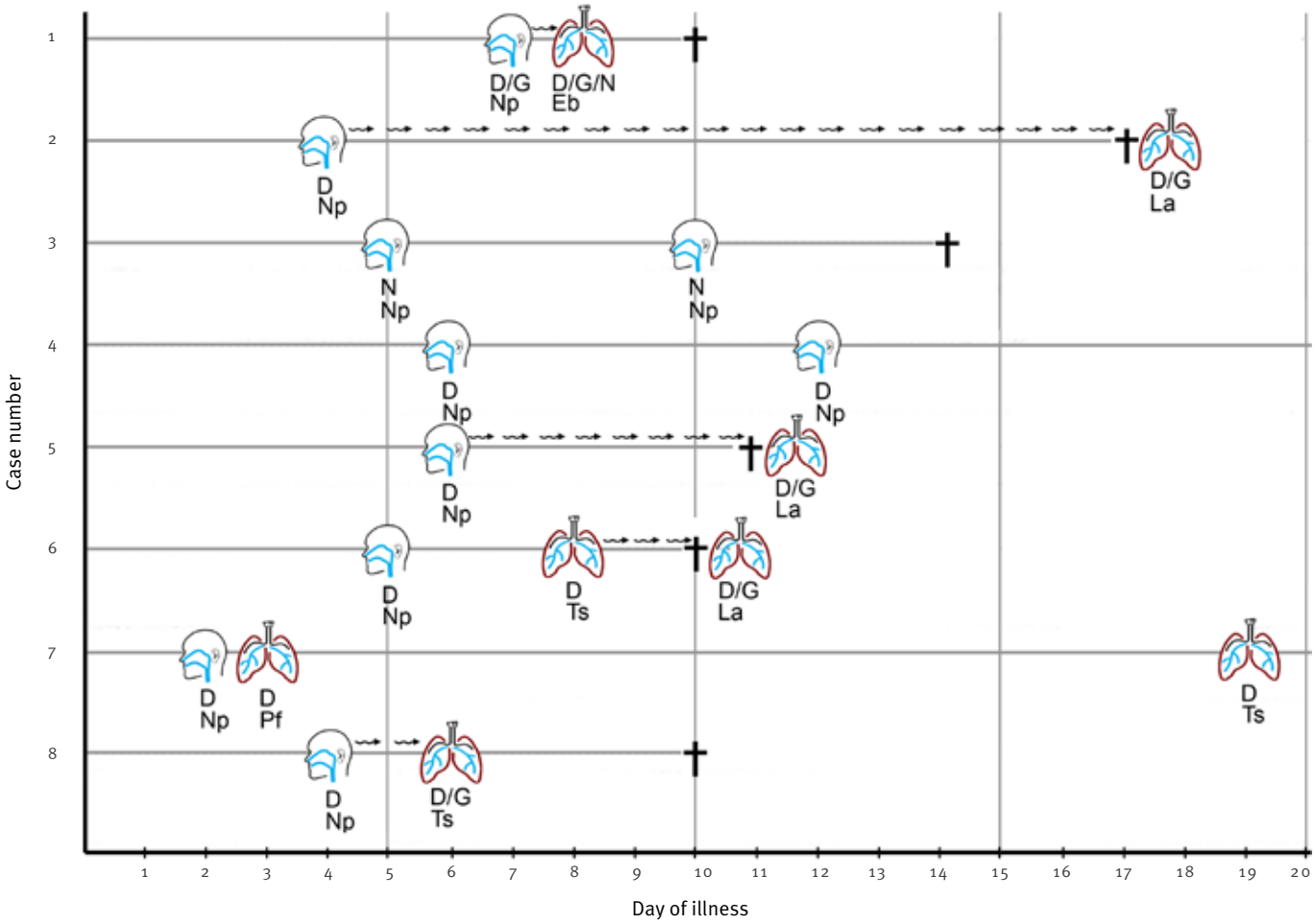
Detection of D222G mutations in the haemagglutinin gene (HA1) using pyro- and Sanger sequencing, influenza A(H1N1) pdm09 cases, Norway 2009/10



Sequence chromatograms at position 222 of influenza A(H1N1)pdm09 virus HA1 gene region derived from four primary clinical samples collected in Norway 2009/10 (4436/09, 3487/09, 3206/09, 4823/09). Chromatograms from pyrosequencing are shown on the left-hand side, those from Sanger sequencing are on the right), illustrating wild-type sequence (panel A), and mutations (panels B–D). In panel D, polymorphism at the 222 codon, with 48% G and 52% A, is illustrated.

Figure 3

Detection of mutations in the haemagglutinin gene (HA1) through the course of illness in influenza A(H1N1)pdm09 cases with serial samples, Norway, 2009/10



Eb	Endobronchial brush
La	Lung autopsy tissue
Np	Nasopharyngeal sample
Pf	Pleural fluid
Ts	Tracheal sample

Upper-airway specimen

Lower-airway specimen

Fatality

Amino acid substitution – D: aspartic acid; G: glycine; N: asparagine.
Genotype changes are shown by arrows.

Of the 13 patients with 222G mutant viruses, we had two or more serial samples available for five (Figure 3). In four of these patients (Cases 2, 5, 6 and 8 in Figure 3), we found only 222D wild type viruses in samples preceding detection of 222G mutant virus. In one patient (Case 1 in Figure 3), both available samples contained 222G mutant virus, but the samples were taken only one day apart, thus reducing the chance of finding differing genotypes. We also had serial samples for three additional non-222G patients (Cases 3, 4 and 7 in Figure 3) who exhibited the same genotype throughout the course of illness (either 222D or 222N).

The date of symptom onset was given for 263 cases: 212 mild, 33 severe non-fatal and 18 fatal. Specimens from the mild cases tended to be collected earlier in the course of illness (mean: 4.3 days after symptom onset; 95% CI: 3.7–4.9) than specimens collected from all severe cases, including fatal cases (mean: 6.5 days; 95% CI: 5.2–7.8) (p=0.001, Mann–Whitney U test). The proportion of 222G mutant viruses was higher in specimens collected late (≥8 days after symptom onset) than in those collected early (<8 days after symptom onset) (p=0.001, Fisher's exact test). Among specimens collected on day 8 or later, more than half were from mild cases (20/39) and there was still a significantly higher proportion of mutant virus in the severe cases, including those that were fatal (6/19 vs 0/20) (p=0.008, Fisher's exact test) (Table 4).

As noted earlier, the median age of the cases included was higher among all the severe cases, including those that were fatal, than the mild cases (p<0.0005, Mann–Whitney U test). The proportion of 222G mutant viruses was higher in specimens collected from patients aged 40 years or older than in younger patients (8.2% (7/85) vs 1.6% (6/371), p=0.004, Fisher's exact test). However, among specimens collected from patients aged 40 years or above, there was still a significantly higher proportion of mutant viruses in the severe cases, including those that were fatal (p<0.0005, Fisher's exact test, data not shown). Thus, the observed difference in 222G prevalence between the different clinical outcome groups is not due to age as a confounding factor.

Discussion

In the present study, we provide further epidemiological evidence of the association between the D222G mutation in HA1 of influenza A(H1N1)pdm09 virus and severe or fatal clinical course. Furthermore, we present evidence that the mutated viruses emerge in individual patients after the onset of illness and demonstrate the presence of mutant virus in both the upper and lower respiratory tract. We also address some potential biases that could conceivably confound the analysis.

The Norwegian cases of infection with HA1 222G genotype viruses have occurred sporadically and do not cluster epidemiologically or in phylogenetic analysis.

As observed by others, the D222G substitution has been observed to occur sporadically in the laboratory when specimens containing predominantly wild-type virus are subjected to virus isolation in eggs or mammalian cell lines [6,13,35].We therefore based our analysis on viral sequences obtained from the primary clinical specimens.

The 222G viruses appear to be rare among circulating strains, but are still quite frequent in patients with severe disease, who are not epidemiologically linked. A likely explanation is that the presence of mutant viruses in these particular individuals experiencing severe disease is due to selective upgrowth of mutant genomes during infection. In all four 222G cases where we could analyse both early and late specimens, we observed a transition from wild-type to mutant virus, which lends support to the hypothesis that the presence of the mutant is due to sporadic emergence rather than widespread circulation.

In our analysis, we could not distinguish between upgrowth from a rare quasispecies, which may be present at inapparent levels in many more cases, and

Table 4

Prevalence of haemagglutinin gene (HA1) 222 genotypes in viruses from influenza A(H1N1)pdm09 cases sampled late in illness (day 8 or later), by clinical outcome, Norway, 2009/10 (n=39)

HA1 222 genotype	Specimens from cases taken on day 8 or later[a]				
	Mild n (%)	Severe n (%)	Fatal n (%)	Severe including fatal n (%)	Total n (%)
222D	17 (85)	8 (73)	3 (38)	11 (58)	28 (72)
222G	0 (0)	3 (27)	3 (38)	6 (32)	6 (15)
222E	3 (15)	0 (0)	1 (13)	1 (5)	4 (10)
222N	0 (0)	0 (0)	1 (13)	1 (5)	1 (3)
Total	**20 (100)**	**11 (100)**	**8 (100)**	**19 (100)**	**39 (100)**

[a] Severe non-fatal and fatal cases are shown separately and jointly.

upgrowth from de novo mutation events. The 222G genotype appears to occur as sporadic mutations with no or little onward transmission, rather than through persistence as a circulating variant. If this is the case, the likelihood of finding mutant viruses is expected to increase during the course of infection. Conceivably, since the severe cases, including those that were fatal, tended to be sampled later in their course of illness than mild cases, the higher occurrence of mutations could alternatively be explained purely as a sampling artefact and not as a marker of virulence. But if the occurrence of the mutation was merely a function of time since infection, there should be no difference in the frequency of 222G mutant in specimens from mild versus severe cases, if taken equally late in infection. Our results show that the significant difference remained even when all early specimens were excluded: thus the difference in sampling day can be ruled out as an alternative explanation for the pattern observed (Table 4).

In several of the cases, simultaneous specimens were available from different body locations. In all of these pairs/sets, the same genotypes were observed, but with differences in the ratio of mutant versus wild type in patients harbouring quasispecies. Other studies have identified D222G quasispecies mainly in endotracheal aspirate or broncheoalveolar lavage (BAL) samples and less frequently in nasopharyngeal aspirate samples [12,14,20]. This discrepancy could be due to the use of different analytical methods. Lower rates of mutant versus wild type in nasal swabs may not have been revealed by conventional sequencing. This was demonstrated by Baldanti et al. [14] in D222G/N cases with available paired nasal swabs and BAL samples, the authors found D222G/N in only one of four cases using direct sequencing, but this proportion increased to three of four cases using clonal analysis. From our study, we cannot exclude the possibility that the occurrence of the mutation as a quasispecies, together with the wild-type, is due to a complementary function of the wild-type. This phenomenon might change if the mutant virus were to develop another fitness-compensatory mutation that permits the mutant to replicate and spread in pure form. This possibility needs to be further studied.

In addition to its possible biological importance the fact that the majority of the cases with 222G mutants carried a mixture of 222D- and 222G-encoding viral genomes, usually with predominance of the wild-type 222D, makes it likely that many influenza A(H1N1) pdm09 cases worldwide carrying D222G mutant viruses could have been missed in analyses that are not sensitive to minor nucleic acid species or where only the majority nucleotide is recorded in the published sequence.

Selleri et al. analysed influenza A(H1N1)pdm09 viral quasispecies and the polymorphism at codon 222 by ultra-deep pyrosequencing [28]. Codon 222 polymorphism was detected in 40.7% of patients by ultra-deep pyrosequencing and in only 7.1% by conventional sequencing. They found that the frequency of polymorphism was significantly higher in samples collected from patients with severe manifestations than in those patients with moderate-mild manifestations. The D222G/N/A mutations were detected as either minor or predominant variants only in severe cases, whereas D222E was equally represented in severe and moderate-mild infections.

The question whether the presence of mutant viruses in lower airways is underestimated due to lack of sampling in the lower respiratory tract was investigated by Baldanti et al. [14] paired nasal swabs and BAL samples from patients admitted to intensive care units for mechanical ventilation or extracorporeal membrane oxygenation were compared with samples from patients with pneumonia not requiring mechanical ventilation and from community patients. By combining data from nasal swabs and BAL samples, the frequency of D222G/N mutants in patients with severe infections increased to 43%, as compared with 7.8% and 0% in patients with moderate and mild infections, respectively [14]. Baldanti et al. also showed that the viral RNA levels were significantly higher in BAL samples than those in nasal swabs [14]. Piralla et al. described the same finding, suggesting higher virus replication in the lower respiratory tract [20].

Watanabe et al. have characterised two of the Norwegian virus isolates included in our study, namely A/Norway/3568/2009 (Norway3568) and A/Norway/3487-2/2009 (Norway3487) [34]. The viruses were isolates from a fatal case (Norway3487) and a mild case (Norway3568): the viruses differ by only 10 amino acids and none of the amino acid changes known to affect virulence were found in PB2, PB1-F2, HA, or NS1, except for an amino acid change to 222G in HA1, in Norway3487. More efficient viral replication in cultured cells and delayed virus clearance from ferret respiratory organs was observed for Norway3487 virus (isolated from a severe case), as compared with Norway3568 (isolated from a mild case). To some extent, Norway3487 virus caused more severe lung damage in non-human primates than did Norway3568 virus. Moreover, the authors found that PB2 derived from Norway3487 contributed to higher polymerase activity, possibly leading to more efficient viral replication in vitro and in vivo, which in turn also could play a role in the increased lung damage.

In our analysis, all cases with detectable 222G mutant viruses had severe illness and the prevalence of 222G increased with the degree of severe outcome. If the conditions for virus growth in lower airways are favourable to 222G mutants and the likelihood of upgrowth of mutant virus increases with duration of infection, it follows that individuals who fail to limit infection to upper airways and fail to eliminate the infection rapidly stand at greater risk of acquiring mutant viruses. Regardless

of virulence, such a mechanism may contribute to the observed pattern of mutants being found primarily in severe cases. However, if the mutant viruses are also more virulent, an increased probability of emergence in patients who already have a long-lasting infection involving the lower respiratory tract probably constitutes a vicious circle phenomenon, which underscores the importance of treatment that helps persons with elevated risk to clear the infection rapidly.

The concept of D222G being a determinant of virulence is supported by a growing body of evidence from in vitro and animal studies including non-human primates [36-45]. Increased tropism of D222G mutant virus for type II pneumocytes [44] may result in more efficient infection, reducing the availability of progenitor cells for essential lung functions and regeneration and thus leading to severe pulmonary impairment. Increased viral titres in the lungs also trigger stronger inflammatory responses, augmenting tissue damage and delaying healing in the infected lungs [44].

While one instance of transmission of 222G mutant virus has been documented [9], the transmitted virus in this case had acquired an additional mutation, (G155E) that may have counteracted the putative 222G-induced receptor-binding changes [46] and there are no indications that effective sustained transmission has taken place. In our data set, one of the patients with mutant viruses was the likely source of infection of a healthcare worker. However, this occurred early in the course of illness: virus from an early specimen, as well as the virus collected from the healthcare worker, was wild type.

In light of the fact that 222G mutant viruses are and remain substantially less transmissible than corresponding wild-type viruses, their immediate public health impact is limited to the individual cases in whom such mutants occur and the fact that they may possibly contribute to the severity of illness. Similar to the apparent situation with avian influenza A(H5N1) virus infection in humans, tropism of 222G mutant viruses for lung cells may contribute to both increased virulence and impaired transmissibility, which may thus be linked traits [47]. On the other hand, when the H275Y neuraminidase mutation conferring oseltamivir resistance in seasonal influenza A(H1N1) viruses was first identified, this mutation also correlated with impaired viral fitness [48], a situation that was abruptly changed with the global emergence of high-fitness resistant viruses during the 2007/08 influenza season [49]. Similarly, further adaption of the newly emerged A(H1N1)pdm09 virus to humans as host species may conceivably lead to compensatory mutations that render the 222G mutants more transmissible, or lead to other changes that influence pathogenicity. The evolution of these viruses therefore needs to be closely monitored in a framework that ensures that virological and clinical data are taken into account.

Acknowledgments

We gratefully acknowledge the essential contributions of primary diagnostic laboratories, clinicians and pathologists in making virus-containing materials and the relevant patient information available to us. We also acknowledge the Department for Infectious Disease Epidemiology, Norwegian Institute of Public Health (NIPH) for invaluable help in supplying the clinical data on many of the fatal and intensive care cases. We would like to thank Jan Oksnes, Department of Bacteriology and Immunology, NIPH, as well as Torstein Aune, Hilde Elshaug, Valentina Johansen, Anne Marie Lund, Grethe Hermansen Krogh, Marianne Morken and Remilyn Ramos-Ocao, Department of Virology, NIPH, for excellent technical assistance.

References

1. Echevarría-Zuno S, Mejía-Aranguré JM, Mar-Obeso AJ, Grajales-Muñiz C, Robles-Pérez E, González-León M, et al. Infection and death from influenza A H1N1 virus in Mexico: a retrospective analysis. Lancet. 2009;374(9707):2072-9.
2. World Health Organization (WHO). Clinical management of human infection with pandemic (H1N1) 2009: revised guidance. Geneva: WHO; November 2009. Available from: http://www.who.int/csr/resources/publications/swineflu/clinical_management/en/index.html
3. van den Brand JM, Stittelaar KJ, van Amerongen G, Rimmelzwaan GF, Simon J, de Wit E, et al. Severity of pneumonia due to new H1N1 influenza virus in ferrets is intermediate between that due to seasonal H1N1 virus and highly pathogenic avian influenza H5N1 virus. J Infect Dis. 2010;201(7):993-9.
4. Stevens J, Blixt O, Glaser L, Taubenberger JK, Palese P, Paulson JC, et al. Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. J Mol Biol. 2006;355(5):1143-55.
5. World Health Organization (WHO). Public health significance of virus mutation detected in Norway. Pandemic (H1N1) 2009 briefing note 17. Geneva: WHO; 20 November 2009. Available from: http://www.who.int/csr/disease/swineflu/notes/briefing_20091120/en/
6. Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A (H1N1) 2009 viruses. Wkly Epidemiol Rec. 2010;85(4):21-2.
7. Kilander A, Rykkvin R, Dudman SG, Hungnes O. Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009-2010. Euro Surveill. 2010;15(9):pii=19498. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19498
8. Antón A, Marcos MA, Martinez MJ, Ramón S, Martínez A, Cardeñosa N, et al. D225G mutation in the hemagglutinin protein found in 3 severe cases of 2009 pandemic influenza A (H1N1) in Spain. Diagn Microbiol Infect Dis. 2010;67(2):207-8.
9. Puzelli S, Facchini M, Spagnolo D, De Marco MA, Calzoletti L, Zanetti A, et al. Transmission of hemagglutinin D222G mutant strain of pandemic (H1N1) 2009 virus. Emerg Infect Dis. 2010;16(5):863-5.
10. Miller RR, MacLean AR, Gunson RN, Carman WF. Occurrence of haemagglutinin mutation D222G in pandemic influenza A(H1N1) infected patients in the West of Scotland, United Kingdom, 2009-10. Euro Surveill. 2010;15(16):pii=19546. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19546
11. Wang B, Dwyer DE, Soedjono M, Shi H, Matlho K, Ratnamohan M, et al. Evidence of the circulation of pandemic influenza (H1N1) 2009 with D222D/G/N/S hemagglutinin polymorphisms during the first wave of the 2009 influenza pandemic. J Clin Virol. 2011;52(4):304-6.
12. Chen H, Wen X, To KK, Wang P, Tse H, Chan JF, et al. Quasispecies of the D225G substitution in the hemagglutinin of pandemic influenza A(H1N1) 2009 virus from patients with severe disease in Hong Kong, China. J Infect Dis. 2010;201(10):1517-21.
13. Mak GC, Au KW, Tai LS, Chuang KC, Cheng KC, Shiu TC, et al. Association of D222G substitution in haemagglutinin of 2009 pandemic influenza A (H1N1) with severe disease. Euro Surveill. 2010;15(14):pii=19534. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=1953
14. Baldanti F, Campanini G, Piralla A, Rovida F, Braschi A, Mojoli F, et al. Severe outcome of influenza A/H1N1/09v infection

associated with 222G/N polymorphisms in the haemagglutinin: a multicentre study. Clin Microbiol Infect. 2011;17(8):1166-9.

15. Melidou A, Gioula G, Exindari M, Chatzidimitriou D, Diza E, Malisiovas N. Molecular and phylogenetic analysis of the haemagglutinin gene of pandemic influenza H1N1 2009 viruses associated with severe and fatal infections. Virus Res. 2010;151(2):192-9.
16. Ikonen N, Haanpää M, Rönkkö E, Lyytikäinen O, Kuusi M, Ruutu P, et al. Genetic diversity of the 2009 pandemic influenza A(H1N1) viruses in Finland. PLoS One. 2010;5(10):e13329.
17. Shikov AN, Sementsova AO, Demina OK, Sergeev AA, Berillo SA, Sergeeva EI, et al. [Genetic variability of isolates of pandemic influenza A virus H1N1 isolated in Russia in 2009]. Mol Gen Mikrobiol Virusol. 2011;(4)23-9. Russian.
18. El Moussi A, Ledesma J, Ben Hadj Kacem MA, Pozo F, Cuevas MT, Hamdoun M, et al. Haemagglutinin D222G mutation found in a fatal case of pandemic (H1N1) flu in Tunisia. Arch Virol. 2012;157(9):1813-4
19. Malato L, Llavador V, Marmier E, Youssef J, Balick WC, Roze H, et al. Pandemic influenza A(H1N1) 2009: molecular characterisation and duration of viral shedding in intensive care patients in Bordeaux, south-west France, May 2009 to January 2010. Euro Surveill. 2011;16(4):pii=19776. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=1977
20. Piralla A, Pariani E, Rovida F, Campanini G, Muzzi A, Emmi V, et al. Segregation of virulent influenza A(H1N1) variants in the lower respiratory tract of critically ill patients during the 2010-2011 seasonal epidemic. PLoS One. 2011;6(12):e28332.
21. Potdar VA, Chadha MS, Jadhav SM, Mullick J, Cherian SS, Mishra AC. Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. PLoS One. 2010;5(3):e9693.
22. Tse H, To KK, Wen X, Chen H, Chan KH, Tsoi HW, et al. Clinical and virological factors associated with viremia in pandemic influenza A/H1N1/2009 virus infection. PLoS One. 2011;6(9):e22534.
23. Houng HS, Garner J, Zhou Y, Lyons A, Kuschner R, Deye G, et al. Emergent 2009 influenza A(H1N1) viruses containing HA D222N mutation associated with severe clinical outcomes in the Americas. J Clin Virol. 2012;53(1):12-5.
24. Yasugi M, Nakamura S, Daidoji T, Kawashita N, Ramadhany R, Yang CS, et al. Frequency of D222G and Q223R hemagglutinin mutants of pandemic (H1N1) 2009 influenza virus in Japan between 2009 and 2010. PLoS One. 2012;7(2):e30946.
25. Nakajima N, Sato Y, Katano H, Hasegawa H, Kumasaka T, Hata S, et al. Histopathological and immunohistochemical findings of 20 autopsy cases with 2009 H1N1 virus infection. Mod Pathol. 2012;25(1):1-13.
26. Alonso M, Rodríguez-Sánchez B, Giannella M, Catalán P, Gayoso J, López Bernaldo de Quirós JC, et al. Resistance and virulence mutations in patients with persistent infection by pandemic 2009 A/H1N1 influenza. J Clin Virol. 2011;50(2):114-8.
27. Drews SJ, Pabbaraju K, Wong S, Tokaryk KL, May-Hadford J, Lee B, et al. Surveillance of autopsy cases for D222G substitutions in haemagglutinin of the pandemic (H1N1) 2009 virus in Alberta, Canada. Clin Microbiol Infect. 2011;17(4):582-4.
28. Selleri M, Piralla A, Rozera G, Giombini E, Bartolini B, Abbate I, et al. Detection of hæmagglutinin D222 polymorphism in influenza A(H1N1)pdm09-infected patients by ultra-deep pyrosequencing. Clin Microbiol infect. 2012 10.1111/j1469-0691.2012.03984.x
29. World Health Organization (WHO). Sequencing primers and protocol. Geneva: WHO; 12 May 2009. Available from: http://www.who.int/csr/resources/publications/swineflu/sequencing_primers/en/index.html
30. World Health Organization (WHO). Protocol No.2: one step conventional RT-PCR for pandemic (H1N1) 2009 HA gene. In: WHO information for laboratory diagnosis of pandemic (H1N1) 2009 virus in humans - revised. Geneva: WHO; 23 November 2009. p. 11. Available from: http://www.who.int/csr/resources/publications/swineflu/WHO_Diagnostic_RecommendationsH1N1_20090521.pdf
31. Lackenby A, Democratis J, Siqueira MM, Zambon MC. Rapid quantitation of neuraminidase inhibitor drug resistance in influenza virus quasispecies. Antivir Ther. 2008;13(6):809-20.
32. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.5c. Seattle: University of Washington; 1993. Available from: http://cmgm.stanford.edu/phylip/
33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731-9.
34. Wilking H, Buda S, von der Lippe E, Altmann D, Krause G, Eckmanns T, et al. Mortality of 2009 pandemic influenza A(H1N1) in Germany. Euro Surveill. 2010;15(49):pii=19741. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19741
35. Ramadhany R, Yasugi M, Nakamura S, Daidoji T, Watanabe Y, Takahashi K, et al. Tropism of pandemic 2009 H1N1 influenza A virus. Front Microbiol. 2012;3:128.
36. Watanabe T, Imai M, Watanabe S, Shinya K, Hatta M, Li C, et al. Characterization in vitro and in vivo of pandemic (H1N1) 2009 viruses isolated from patients. J Virol. 2012;86(17):9361-8.
37. Abed Y, Pizzorno A, Hamelin ME, Leung A, Joubert P, Couture C, et al. The 2009 pandemic H1N1 D222G hemagglutinin mutation alters receptor specificity and increases virulence in mice but not in ferrets. J Infect Dis. 2011;204(7):1008-16.
38. Belser JA, Jayaraman A, Raman R, Pappas C, Zeng H, Cox NJ, et al. Effect of D222G mutation in the hemagglutinin protein on receptor binding, pathogenesis and transmissibility of the 2009 pandemic H1N1 influenza virus. PLoS One. 2011;6(9):e25091.
39. Chutinimitkul S, Herfst S, Steel J, Lowen AC, Ye J, van van Riel D, et al. Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. J Virol. 2010;84(22):11802-13.
40. Ilyushina NA, Khalenkov AM, Seiler JP, Forrest HL, Bovin NV, Marjuki H, et al. Adaptation of pandemic H1N1 influenza viruses in mice. J Virol. 2010;84(17):8607-16.
41. Liu Y, Childs RA, Matrosovich T, Wharton S, Palma AS, Chai W, et al. Altered receptor specificity and cell tropism of D222G hemagglutinin mutants isolated from fatal cases of pandemic A(H1N1) 2009 influenza virus. J Virol. 2010;84(22):12069-74.
42. Memoli MJ, Bristol T, Proudfoot KE, Davis AS, Dunham EJ, Taubenberger JK. In vivo evaluation of pathogenicity and transmissibility of influenza A(H1N1)pdm09 hemagglutinin receptor binding domain 222 intrahost variants isolated from a single immunocompromised patient. Virology. 2012;428(1):21-9.
43. Seyer R, Hrincius ER, Ritzel D, Abt M, Mellmann A, Marjuki H, et al. Synergistic adaptive mutations in the hemagglutinin and polymerase acidic protein lead to increased virulence of pandemic 2009 H1N1 influenza A virus in mice. J Infect Dis. 2012;205(2):262-71.
44. Watanabe T, Shinya K, Watanabe S, Imai M, Hatta M, Li C, et al. Avian-type receptor-binding ability can increase influenza virus pathogenicity in macaques. J Virol. 2011;85(24):13195-203.
45. Ye J, Sorrell EM, Cai Y, Shao H, Xu K, Pena L, et al. Variations in the hemagglutinin of the 2009 H1N1 pandemic virus: potential for strains with altered virulence phenotype? PLoS Pathog. 2010;6(10):e1001145.
46. Takahashi T, Hashimoto A, Maruyama M, Ishida H, Kiso M, Kawaoka Y, et al. Identification of amino acid residues of influenza A virus H3 HA contributing to the recognition of molecular species of sialic acid. FEBS Lett. 2009;583(19):3171-4.
47. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. Avian flu: influenza virus receptors in the human airway. Nature. 2006;440(7083):435-6.
48. Ives JA, Carr JA, Mendel DB, Tai CY, Lambkin R, Kelly L, et al. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. Antiviral Res. 2002;55(2):307-17.
49. Meijer A, Lackenby A, Hungnes O, Lina B, van der Werf S, Schweiger B, et al. Oseltamivir-resistant influenza virus A (H1N1), Europe, 2007-08 season. Emerg Infect Dis. 2009 Apr;15(4):552-60.

Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrug-resistant clone

S A Chisholm (stephanie.chisholm@hpa.org.uk)[1], M Unemo[2], N Quaye[1], E Johansson[2], M J Cole[1], C A Ison[1], M JW Van de Laar[3]
1. Sexually Transmitted Bacteria Reference Laboratory, Health Protection Agency, Colindale, London, United Kingdom
2. National Reference Laboratory for Pathogenic Neisseria, Department of Laboratory Medicine, Clinical Microbiology, Örebro University Hospital, Örebro, Sweden
3. European Centre for Disease Prevention and Control, Stockholm, Sweden

Citation style for this article:
Chisholm SA, Unemo M, Quaye N, Johansson E, Cole MJ, Ison CA, Van de Laar MJ. Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrug-resistant clone. Euro Surveill. 2013;18(3):pii=20358. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20358

Article submited on 31 August 2012 / published on 17 January 2013

Treatment of gonorrhoea is threatened by antimicrobial resistance, and decreased susceptibility and resistance to recommended therapies is emerging in Europe. Current associations between resistance and molecular type remain poorly understood. Gonococcal isolates (n=1,066) collected for the 2009 and 2010 European Gonococcal Antimicrobial Surveillance Programme were typed by *Neisseria gonorrhoeae* multi-antigen sequence typing (NG-MAST). A total of 406 sequence types (STs) were identified, 125 of which occurred in ≥two isolates. Seven major genogroups of closely related STs (varying by ≤1% at just one of the two target loci) were defined. Genogroup 1407 (G1407), observed in 20/21 countries and predominant in 13/21 countries, accounted for 23% of all isolates and was associated with decreased susceptibility to cefixime and resistance to ciprofloxacin and raised minimum inhibitory concentrations for ceftriaxone and azithromycin. Genogroup 225 (G225), associated with ciprofloxacin resistance, was observed in 10% of isolates from 19/21 countries. None of the other genogroups were associated with antimicrobial resistance. The predominance of a multidrug-resistant clone (G1407) in Europe is worrying given the recent reports of recommended third generation cephalosporins failing to treat infections with this clone. Identifying associations between ST and antimicrobial resistance aids the understanding of the dissemination of resistant clones within a population and could facilitate development of targeted intervention strategies.

Introduction

Gonorrhoea continues to present a public health problem within the European Union/European Economic Area (EU/EEA). In 2010, a total of 32,028 cases of gonorrhoea were reported by 28 European countries, giving a rate of 10.4 per 100,000 population [1]. Given the serious implications of gonococcal infection, which include pelvic inflammatory disease, infertility, ectopic pregnancy, early abortion and a fivefold increase in risk of human immunodeficiency virus (HIV) transmission [2], the need for effective treatment and control of gonorrhoea is evident. A significant challenge to this is that during the antibiotic era, the aetiological agent, *Neisseria gonorrhoeae*, has consistently developed resistance to each antibiotic introduced for widespread empirical treatment [3].

Empirical treatment of gonorrhoea is based on national or regional guidelines, many of which currently recommend third generation cephalosporins as first line therapy. Up until late in 2012, guidelines for Europe recommend 250 mg of ceftriaxone as a single intramuscular dose or 400 mg of cefixime as a single oral dose [4]. These were recently revised to recommend 500mg ceftriaxone combined with 2g azithromycin [5] in response to growing evidence that cephalosporin therapy is becoming compromised by the emergence of isolates exhibiting decreased susceptibility and resistance to the third generation cephalosporins, which have now caused treatment failure in several EU/EEA countries [6-11]. In 2010, the European gonococcal antimicrobial susceptibility programme (Euro-GASP) demonstrated that 9% of isolates from 21 European countries showed decreased susceptibility to cefixime as well as an increasing trend in minimum inhibitory concentrations (MICs) for ceftriaxone, and a high prevalence of resistance to ciprofloxacin (53%) and azithromycin (7%) [12]. In the absence of any alternative antimicrobial treatments, there is clearly an urgent need to develop strategies to limit dissemination of resistant strains to maintain effective treatment.

Molecular epidemiological surveillance has the capacity to provide novel information on the emergence and dissemination of antimicrobial resistant gonococcal

clones which might facilitate appropriate patient management and targeted intervention strategies. *Neisseria gonorrhoeae* multi-antigen sequence typing (NG-MAST) is highly discriminatory, reproducible, objective and transferrable and is well suited to applications aiming to answer short term epidemiological questions. Several studies have applied NG-MAST to characterise gonococci exhibiting decreased susceptibility and resistance to various antimicrobials including the third generation cephalosporins [13-21]. However molecular epidemiological surveillance of consecutive isolates of all antimicrobial phenotypes within a population is necessary to explore associations between genotype and antimicrobial resistance profile and to fully understand how resistant strains emerge and disseminate. While several such surveys have been conducted within individual countries in the EU/EEA [22-25] and elsewhere [26-28], most examined gonococcal populations had been collected in 2008 or earlier and, accordingly, associations between sequence type (ST) and antimicrobial susceptibility may no longer be valid. Furthermore examples of decreased susceptibility to recommended third generation cephalosporins were in most cases not circulating in the gonococcal population examined at the time of sampling.

The current study aimed to assess the public health benefit of NG-MAST molecular epidemiological typing of gonococcal isolates as part of Euro-GASP.

Methods

Isolate collection

N. gonorrhoeae isolates were selected from the most recent sampling period (2010) within Euro-GASP. The sampling strategy for this multi-centre sentinel surveillance scheme has been described previously [12]. Briefly, countries are requested to submit 55 consecutive gonococcal isolates twice annually (in May/June and November/December) aiming as far as possible to provide isolates representative of the national distribution of cases of gonorrhoea. Where available, countries submitted information on date specimen obtained, specimen site, sex of patient, age, sexual orientation, previously diagnosed with gonorrhoea, and concurrent sexually transmitted infection (STI) diagnosed in this episode.

A total of 1,066 isolates were selected for molecular epidemiological typing from 21 EU/EEA countries. The criteria for selection were that isolates were consecutive to ensure no bias in antimicrobial susceptibility, and that a full antimicrobial resistance profile had been determined as part of EURo-GASP.

Most (n=828; 78%) were solely from the first sampling period in May/June 2010 in Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Sweden and the United Kingdom (UK). Isolates were selected from this sampling period for all other participating countries but further isolates (n=139; 13%) were additionally selected from the second sampling period (November/December 2010) for countries submitting low numbers (‹50 isolates in the first collection period): Cyprus, Greece, Hungary, Ireland, Romania and Slovakia. The remaining 100 isolates (9%) were additionally selected from 2009 for countries submitting very low numbers (‹50 isolates from both sampling periods in 2010) (Latvia, Malta, Slovenia) or for countries where isolates from the second sampling period in 2010 were unavailable and ‹50 isolates had been submitted in the first sampling period of 2010 (Portugal, Norway and Spain). As Cyprus, Hungary and Romania did not participate in EURO-GASP 2009, only very low numbers (≤17) were available from 2010 for inclusion in the current study.

Antimicrobial susceptibility testing

Gonococcal isolates from each country were submitted on Microbank beads (Pro-Lab Diagnostics, Cheshire, UK) frozen on dry ice to one of three centres (the Health Protection Agency, UK, Örebro University Hospital, Sweden and the Statens Serum Institut, Denmark) for centralised susceptibility testing. All isolates were tested for susceptibility to ciprofloxacin, azithromycin and spectinomycin by a previously decribed breakpoint method [12]. MICs were determined by agar dilution for gentamicin as described [12,29] and by Etest for cefixime and ceftriaxone (AB Biomerieux, Solna, Sweden). The MIC was confirmed by Etest for any isolates resistant to azithromycin by the breakpoint method or showing gentamicin MICs›8 mg/L. All isolates were also tested for penicillinase production using the chromogenic reagent Nitrocefin (Oxoid, Basingstoke, UK).

Determination and analysis of *Neisseria gonorrhoeae* multi-antigen sequence types

Permission to type isolates submitted for susceptibility testing was obtained from all participating laboratories. All typing was performed at either the Health Protection Agency, UK, or at Örebro University Hospital, Sweden. Isolates originally referred to the Statens Serum Institute in Denmark for susceptibility testing were sent frozen on Microbank beads to the Health Protection Agency, London, UK for typing. A total of 1,066 isolates were typed as described [30] by NG-MAST, which differentiates strains on the basis of sequence variation in fragments of two hypervariable genes, the porin PorB (*porB*) gene and subunit B of the transferrin binding protein (*tbpB*) gene. Allele numbers and STs, including any new alleles or STs were assigned via the online NG-MAST database (www.ng-mast.net).

The alleles defining the most frequently observed STs (represented by 10 isolates or more), were compared against all alleles in the dataset to determine if predominant STs were part of a wider collection of closely related types. Similarity of alleles was evaluated in BioNumerics v6.1 by neighbour joining multiple alignment, followed by individual pairwise alignment against the most frequent allele to determine number of base pair (bp) differences. For example, for ST1407 (*porB* 908, *tbpB* 110), all *porB* alleles from isolates

Table 1

Most frequently observed *Neisseria gonorrhoeae* multi-antigen sequence types in each country, and frequency of the three most common among all isolates (n=1,066), EU/EEA countries, 2009–2010

Country	Most frequent ST	STs n	ST1407 n (%)	ST2992 n (%)	ST225 n (%)	Isolates typed n
Austria	ST1407	20	16 (32)	2 (4)	0 (0)	50
Belgium	ST1407	31	7 (14)	4 (8)	1 (2)	50
Cyprus	ST3128	5	0 (0)	0 (0)	0 (0)	12
Denmark	ST225/3158	36	3 (6)	1 (2)	6 (12)	50
France	ST2/2992	27	4 (8)	7 (14)	2 (4)	50
Germany	ST25	18	14 (28)	0 (0)	0 (0)	50
Greece	ST5405[a]/5505[a]	24	6 (12)	0 (0)	1 (2)	50
Hungary	ST5332[a]	10	2 (12)	0 (0)	0 (0)	17
Ireland	ST2992	25	3 (6)	16 (32)	0 (0)	50
Italy	ST1407	25	17 (34)	5 (10)	0 (0)	50
Latvia	ST3227	14	0 (0)	0 (0)	2 (7)	29
Malta	ST225	16	3 (6)	1 (2)	23 (46)	50
Netherlands	ST1407	52	16 (16)	7 (7)	3 (3)	100
Norway	ST2992	26	7 (14)	11 (22)	0 (0)	49
Portugal	ST1407	28	7 (14)	4 (8)	0 (0)	50
Romania	ST1407/4120	7	2 (22)	0 (0)	0 (0)	9
Slovakia	ST437	19	2 (4)	0 (0)	0 (0)	50
Slovenia	ST1407	19	14 (28)	2 (4)	8 (16)	50
Spain	ST1407	43	28 (28)	9 (9)	0 (0)	100
Sweden	ST225	31	1 (2)	3 (6)	4 (8)	50
United Kingdom	ST1407	62	14 (14)	4 (4)	0 (0)	100
Total N (%)			**166 (16)**	**76 (7)**	**50 (5)**	**1,066 (100)**

EEA: European Economic Area; EU: European Union; ST: sequence type.

[a] New ST.

with *tbpB* allele 110 were compared for similarity to *porB* allele 908, and all *tbpB* alleles from isolates containing *porB* allele 908 were compared for similarity to *tbpB* allele 110. Different STs were assigned to a 'genogroup', if one identical allele was shared and the other allele showed ≥99% similarity (≤5 bp difference for *porB* and ≤4 bp for *tbpB*). Genogroups were named using the letter G followed by the number of the predominant ST within each group. For example G1407 is the genogroup in which ST1407 is the predominant ST.

Statistical analyses

Potential associations between genogroups and (i) antimicrobial susceptibilities and (ii) patient characteristics (sex, age and sexual orientation) were explored firstly by univariate analysis and then where appropriate by multivariate analysis using STATA v11.2. For this analysis, patients were subdivided into age groups (0–24 years, 25–34 years, 35–44 years or ≥45 years). Most of the patients in the 0–24 year age group were aged 15–24 years (99%; 360/363), while the remaining three patients were aged 0–7 years.

For the univariate analysis, crude odds ratios (OR) and 95% confidence intervals (CI) were calculated where datasets contained sufficient numbers. A Pearson chi-squared test was used to test if these ORs were significantly different from one (i.e. testing the null hypothesis that there was no difference in odds of resistance/decreased susceptibility between the group in question and the specified baseline group). For datasets where a cell equalled zero, this analysis could not be performed. In these cases of small cell numbers, Fisher's exact test was performed.

The multivariate analysis used logistic regression to model the odds of associations between genogroup and resistance controlling for other variables. The P value produced from the Wald test was used to test the null hypothesis that the odds ratios were not significantly different from one.

Results

Frequency of *Neisseria gonorrhoeae* multi-antigen sequence types

Of the 1,066 isolates typed, 406 different STs were identified, representing 313 *porB* alleles and 113 *tbpB* alleles. One hundred and twenty five clusters (≥2 isolates with the same ST) and 281 STs represented by a single isolate were identified, and 216 new STs defined. The most frequently observed types (represented by ≥10 isolates) were STs 1407 (n=166), 2992 (n=76), 225 (n=50), 25 (n=18), 2 (n=17), 359 (n=16), 387(n=15), 437 (n=14), 3227 (n=12), 5405 (n=10) and 5595 (n=10), with 5405 and 5595 being new STs. Of these, all had unique *porB* alleles but some STs shared the same *tbpB* allele (STs 1407 and 5595, STs 225 and 437, STs 2292 and 359 and STs 387 and 3227 shared, respectively, *tbpB* alleles 110, 4, 29 and 118).

The prevalence of the three most common STs (1407, 2992 and 225) varied between countries (Table 1). ST1407 accounted for >10% of isolates in 13/21 countries, whereas ST2992 and ST225 were observed in >10% of isolates in 3/21 countries in both cases (Table 1). ST1407 was the most frequent ST observed in Austria, Belgium, Italy, the Netherlands, Portugal, Romania, Slovenia, Spain and the UK (Table 1).

Definition and frequency of *Neisseria gonorrhoeae* multi-antigen sequence typing genogroups

Further analysis of the *porB* and *tbpB* alleles of the STs observed in ≥10 isolates showed that several other STs within the total collection were highly related to these, sharing one identical allele and differing by ≤1% at the other allele (in most cases, the *porB* allele). Seven major genogroups, characterised as G1407, G225, G2992, G25, G387, G359 and G2, were defined (Table 2), encompassing 557 (52%) of isolates. The remaining 509 isolates were represented by 313 STs, of which only STs 5405 (n=10), 995 (n=9), 951 (n=8), 4347 (n=8), 5505 (n=8), 292 (n=7), 2400 (n=7), 5402 (n=7), 384 (n=6), 1034 (n=6),1780 (n=6) and 5598 (n=5) were observed in ≥5 isolates. The largest and most diverse genogroups were G1407, comprising ST1407 and 25 STs differing by ≤5 bp in the *porB* allele and two STs by 1 bp in *tbpB*, and G225 comprising ST225 and 28 other STs differing by ≤5 bp at the *porB* locus (Table 2).

Distribution of *Neisseria gonorrhoeae* multi-antigen sequence typing genogroups

G1407 was observed in all countries examined, with the exception of Latvia, and was the most frequent type, observed in 15–50% of isolates from Austria, Belgium, Cyprus, Denmark, England and Wales, Greece, Hungary, Italy, Portugal, Romania, Slovakia, Slovenia, Spain and the Netherlands (Figure). In contrast G1407 was comparatively uncommon (≤8%) in France, Ireland, Malta and Sweden (Figure). G225 was observed in 19/21 countries and was the most frequently observed type in Malta and Sweden and was most common after

Table 2

Neisseria gonorrhoeae multi-antigen sequence types within the seven major genogroups defined based on sequence similarity at *porB* and *tbpB* alleles, EU/EEA countries, 2009–2010

Genogroup n (%)[a]	Predominant ST (n)	STs within each genogroup (n)	
		≥99% similar to *porB* allele x[b]	≥99% similar to *tbpB* allele x[c]
G1407 248 (23)	1407 (166)	5595 (10), 3149 (9), 3158 (8), 3128 (7), 4120 (7), 5570 (5), 5600 (4), 2212 (3), 5594 (3), 3431 (1), 3779 (1), 4275 (1), 4359 (1), 4951 (1), 5588 (1), 5619 (1), 5622 (1), 4974 (1), 5480 (1), 5625 (1), 5581 (1), 5649 (1), 5205 (1), 5599 (1), 5591 (1)	5332 (6), 4741 (4)
G225 107 (10)	225 (50)	437 (14), 5463 (6), 1132 (5), 5017 (2), 5423 (2), 2616 (2), 3150 (2), 5180 (2), 205 (1), 346 (1), 1342 (1), 1365 (1), 1399 (1), 2687 (1), 289 (1), 1340 (1), 2202 (1), 3141 (1), 3153 (1), 3952 (1), 4315 (1), 5703 (1), 2625 (1), 5655 (1), 3056 (1), 4473 (1), 5374 (1), 5330 (1)	
G2992 84 (8)	2992 (76)	5515 (2), 5049 (1) 5194 (1) 5237 (1), 5192 (1) 5385 (1)	5227 (1)
G25 43 (4)	25 (18)	51 (7), 3003 (4), 356 (3), 4589 (2), 5384 (2), 273 (1), 881 (1), 5424 (1), 807 (1), 5341 (1), 4338 (1), 5172 (1)	
G387 37 (3)	387 (15)	3227 (12), 5185 (4), 5191 (1), 5486 (1), 5503 (1), 5186 (1), 5190 (1), 5498 (1)	
G359 20 (2)	359 (16)	1929 (2), 5485 (1), 1313 (1)	
G2 18 (2)	2 (17)	226 (1)	

EEA: European Economic Area; EU: European Union; *porB*: porin B; ST: sequence type; *tbpB*: transferrin binding protein subunit B.

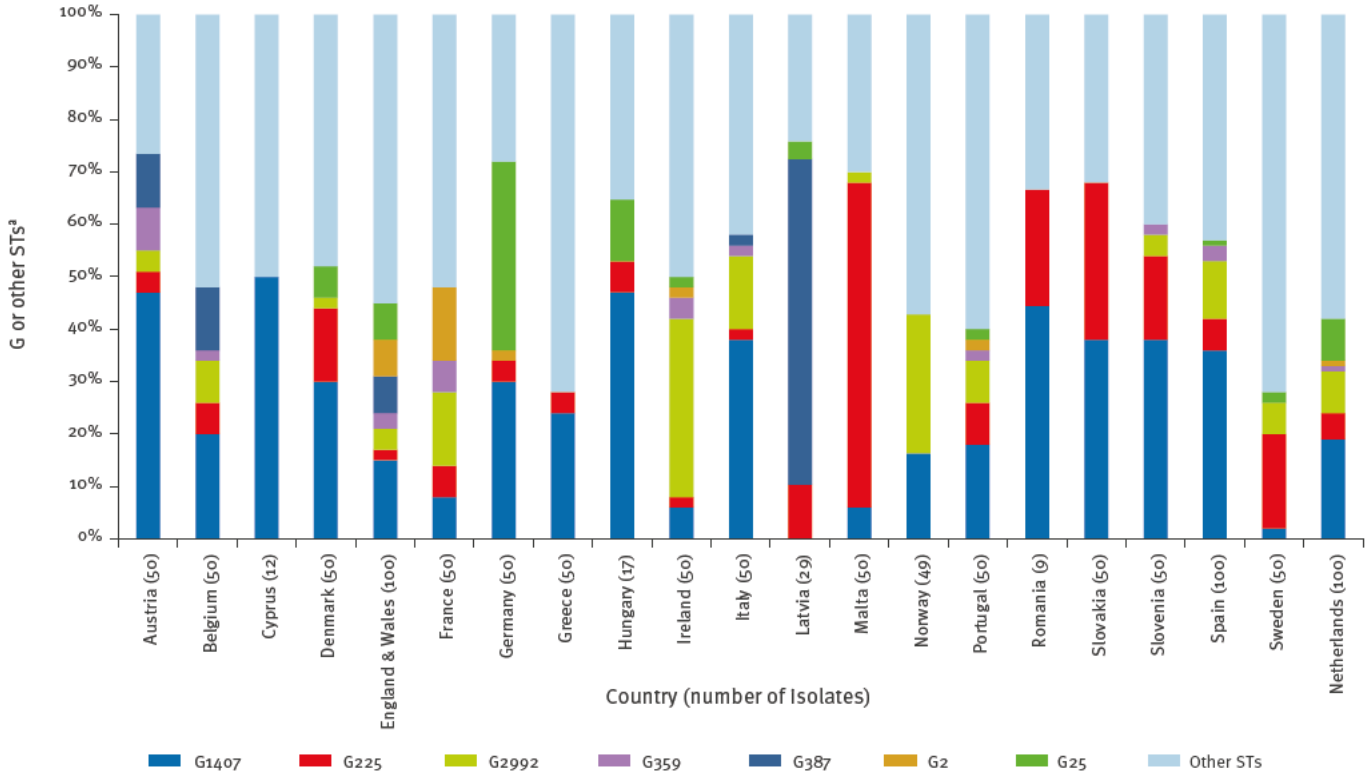
[a] The percentages are calculated from a total of 1,066 isolates.

[b] STs that share an identical *tbpB* allele with the most frequent ST within the genogroup but vary at the *porB* allele.

[c] STs that share an identical *porB* allele with the most frequent ST within the genogroup but vary at the *tbpB* allele.

Figure

Proportion of various *Neisseria gonorrhoeae* multi-antigen sequence typing genogroups within each participating country, EU/EEA countries, 2009–2010



EEA: European Economic Area; EU: European Union; G:genogroup; ST: sequence type.

[a] Other STs refer to the isolates with STs which did not belong to a defined G.

Table 3

Characteristics of patients infected with the most frequently observed *Neisseria gonorrhoeae* multi-antigen sequence typing genogroups/sequence type, EU/EEA countries, 2009–2010

Genogroup (n)	Mean patient age (range in years)	Patients of known sex n	Male patients n (%)	Patients of known sexual orientation[a] n	Males of known sexual orientation n	MSM n (%[b])	Heterosexuals of both sex n (%[b])
G1407 (248)	33.3 (17–69)	244	218 (89)	155	129	72 (46)	83 (54)
G225 (107)	29.1 (7–56)	104	85 (82)	78	59	13 (17)	65 (83)
G2992 (84)	31.7 (18–66)	84	81 (96)	41	38	36 (88)	5 (12)
G25 (43)	24.6 (16–40)	43	18 (42)	35	10	2 (6)	33 (94)
G387 (37)	28.6 (18–69)	36	21 (58)	30	15	0 (0)	30 (100)
G359 (20)	32.9 (18–69)	20	16 (80)	11	7	5 (45)	6 (55)
G2 (18)	22.1 (15–38)	18	12 (67)	11	5	0 (0)	11 (100)
ST5405[c] (10)	30.9 (6–51)	10	8 (80)	9	7	0 (0)	9 (100)
All isolates (1,066)	31.4 (0–76)	1,050	868 (83)	660	478	222 (34)	438 (66)

EEA: European Economic Area; EU: European Union; MSM: men who have sex with men.

[a] All female patients were included in the heterosexual patient group.
[b] The percentage is relative to the number of patients of known sexual orientation.
[c] ST5405 was not found to be part of a larger genogroup.

G1407 in Denmark, Romania, Slovakia and Slovenia, accounting for 14–30% of isolates (Figure). G2992 was observed in 14/21 countries and predominated in Ireland and Norway, while G25 predominated in Germany (Figure).

Neisseria gonorrhoeae multi-antigen sequence typing genogroups and epidemiological characteristics of linked patients

The most frequently collected variables were age and sex, known for 1,030 and 1,050 patients, respectively. Sexual orientation was known for only 586 patients (478 males and 108 females) (Table 3).

Statistical analysis showed there was no clear association with age category for most genogroups, with the exception of G1407 which was significantly less common in patients aged <25 years (OR: 0.43; p<0.0001, chi-squared test), whereas G2 was significantly more common in this younger group (OR: 15.63; p=0.0004, chi-squared test). The proportion of males infected was significantly lower for G25 (OR: 0.13; p<0.0001, chi-squared test) and G387 (OR: 0.27; p=0.0001, chi-squared test) but was significantly higher for G2992 (OR: 6.14; p=0.0005, chi-squared test) and G1407 (OR: 2.01; p=0.0017, chi-squared test). Where sexual orientation was known, statistical associations were observed between men who have sex with men (MSM) and infection with G2992 (OR: 15.86; p<0.0001, chi-squared test) and G1407 (OR: 1.75; p=0.003, chi-squared test), although 54% of patients infected with G1407 were heterosexual (Table 3). MSM were less frequently infected with G225 (OR: 0.039; p=0.0015, chi-squared test), G387 (p<0.0001, Fisher's exact test) and G2 (p=0.01, Fisher's exact test).

Neisseria gonorrhoeae multi-antigen sequence typing genogroups or sequence types and antimicrobial resistance

The consensus antimicrobial susceptibility category or modal MIC was calculated for all genogroups/STs represented by ≥10 isolates (Table 4). Most deviation from the consensus was observed for azithromycin, which supplementary testing by Etest showed was due to isolates exhibiting MICs close to the breakpoint for resistance (1.0 mg/L).

The modal cefixime MIC was raised for G1407 compared with other genogroups (Table 4), with 77/248 isolates (31%) showing decreased susceptibility to cefixime (MIC >0.125 mg/L) while 176/248 isolates (71%) had MICs ≥0.125mg/L. Almost all isolates (239/248; 96%) had cefixime MICs ≥0.06mg/L. The association between G1407 and decreased susceptibility to cefixime was highly significant (OR: 17.9; p<0.0001, chi-squared test) by univariate analysis and by multivariate analysis to control for patient age, sex and sexual orientation (OR: 31.2; p<0.0001, chi-squared test). There were no examples of decreased ceftriaxone susceptibility (MIC >0.125 mg/L), but the modal MIC was raised for G1407 compared with other genogroups (Table 4). All isolates of G1407 were ciprofloxacin resistant (p<0.0001, Fisher's exact test). While the consensus azithromycin susceptibility for G1407 was sensitive, the proportion of isolates showing MICs just above the breakpoint and so in the resistant category was significant after controlling for patients age, sex and sexual orientation (OR: 11.5; p<0.0001, chi-squared test). Decreased cefixime susceptibility was observed in a further nine STs representing 10 isolates in the total study population. Of these, seven isolates shared *tbpB* allele 110 but the six *porB* alleles differed from allele 908 by >5 bp. One isolate possessed *porB* allele 908 but differed markedly at the *tbpB* allele from 110. The STs of the remaining two isolates (STs 3168 and 5606) appeared unrelated to ST1407.

All isolates belonging to G225 were resistant to ciprofloxacin, indicating a statistically significant association by Fisher's exact test (p<0.0001). All ten isolates of ST5405 were resistant to ciprofloxacin also (Table 4), significant association by Fisher's exact test (p=0.0032).

Discussion

To our knowledge this is the first study to supplement a regional antimicrobial resistance surveillance programme with molecular typing to provide enhanced surveillance of gonorrhoea. This approach has not only provided novel information on NG-MAST STs circulating in 21 of the 30 EU/EEA countries, but also provides insight into the dissemination of antimicrobial resistant gonococcal clones within the region.

While considerable diversity of gonococcal STs exists both within and between countries, some STs predominate, and this is even more apparent if STs are grouped into genogroups to facilitate robust statistical analyses. Most notably, G1407 accounted for 23% of isolates overall, and predominated in 13/21 countries. Relatedness of NG-MAST STs has been considered in comparatively few studies and multiple approaches are described [14,15,21,25,31,32]. Failure to consider relatedness of STs could lead to the prevalence of clonal groups being underestimated and reduces the power of any statistical analyses exploring associations between organism and/or patient characteristics and STs. Our approach was to consider strains highly related if strains varied at just one allele by ≤1%. Use of such a high cut-off ensures only very closely related strains are clustered, but given the variation in approaches described thus far, there is a need to agree a standard approach in the near future to allow comparability between studies.

Most molecular surveys of consecutive gonococcal isolates from earlier sampling periods ranging between 2003 to 2005 did not demonstrate predominance of ST1407 or related types in individual European countries [24,25,33,34] or elsewhere [26,28]. In contrast ST1407 was documented in more recent surveys in Portugal [25] and Norway [23], and a centre that

Table 4

Consensus antimicrobial susceptibility results for *Neisseria gonorrhoeae* multi-antigen sequence typing genogroups/sequence type represented by 10 or more isolates, EU/EEA countries, 2009–2010

Genogroup or ST	Isolates n	Beta-lactamase	Consensus resistance category (Isolates differing from the consensus)			Modal MIC mg/L (range)		
			Azithromycin	Ciprofloxacin	Spectinomycin	Gentamicin	Cefixime[a]	Ceftriaxone[a]
G1407	248	Negative	S (44[b])	R (0)	S (0)	8 (4–16)	0.125 (0.032–0.25)	0.047 (0.008–0.125)
G225	107	Negative	S (1[c])	R (0)	S (0)	8 (4–16)	0.023 (‹0.016–0.064)	0.016 (0.003–0.047)
G2992	84	Negative	S (14[d])	S (1)	S (0)	8 (2–8)	0.023 (‹0.016–0.064)	0.006 (0.002–0.016)
G25	43	Negative	S (0)	S (1)	S (0)	8 (4–8)	‹0.016 (‹0.016–0.064)	0.003 (‹0.002–0.016)
G387	37	Negative	S (0)	S (0)	S (0)	8 (4–8)	‹0.016 (‹0.016–0.016)	‹0.002 (‹0.002–0.002)
G359	20	Negative	S (1[c])	S (0)	S (0)	8 (4–16)	0.064 (‹0.016–0064)	0.008/0.012 (0.004–0.023)
G2	18	Negative	S (0)	S (0)	S (0)	4 (2–8)	‹0.016 (‹0.016–0.032)	0.003 (‹0.002–0.012)
ST5405	10	Negative	S (0)	R (0)	S (0)	8 (4–8)	‹0.016 (‹0.016–0.023)	0.004/0.006 (0.003–0.006)

EEA: European Economic Area; EU: European Union; MIC: minimum inhibitory concentration; S: susceptible; ST: sequence type; R: resistant.

[a] Cefixime and Ceftriaxone MICs are described according to the Etest scale.
[b] 40 of the 44 with an R category were within one doubling dilution of the breakpoint.
[c] Isolates that differ from the consensus were within one doubling dilution of the breakpoint.
[d] 10 of the 14 with an R category were within one doubling dilution of the breakpoint.

performs molecular typing of all cases of gonorrhoea in Scotland showed that ST1407 first emerged in 2007 and by 2009 accounted for 15.4% of cases [35]. The current study provides further evidence that the wide dissemination of ST1407 in the EU is likely to be a recent phenomenon. ST1407 and related STs are known to be distributed globally, documented in studies specifically investigating decreased susceptibility and resistance to third generation cephalosporins, in Europe [6,9,13,14,19], the United States [20], Canada [21], Australia [18], Hong Kong [17] and Taiwan [36] and in a survey conducted in Japan [27]. ST1407 has also caused most of the treatment failures with third generation cephalosporins in EU/EEA countries [6-11].

Our study demonstrates that there appear to be no isolates of G1407 that are highly sensitive to cefixime, with 96% of isolates showing MICs ≥0.06mg/L, and presents the first clear evidence that all G1407 isolates circulating currently in the general gonococcal population in the EU/EEA are ciprofloxacin resistant, strongly associated with decreased susceptibility/resistance to cefixime and show raised MICs for ceftriaxone and azithromycin. This highlights the potential of molecular epidemiological typing as a tool to predict antimicrobial resistance, as it is evident that ciprofloxacin and cefixime in particular would be inappropriate regimes to treat G1407 infection. Two isolates were identified that exhibited decreased susceptibility to cefixime but were considered unrelated to G1407. The low incidence of these types may suggest they represent less biologically fit gonococcal clones and/or have recently been imported from regions such as the Far East, where greater diversity in STs showing cefixime decreased susceptibility is reported [15-17,36]. ST1407 is known to be part of a major globally predominant clone, multilocus sequence type (MLST) ST1901 which probably originated in Japan [11], and recent analysis of ten additional genomic markers has provided further evidence of the clonal nature of ST1407 and closely related NG-MAST STs [32]. MLST ST7363 is also reported as a major clone associated with decreased susceptibility and resistance to cephalosporins, which also has shown its ability to develop high-level resistance to cefixime and ceftriaxone [37]. MLST was not performed in the current study but is an excellent means of determining longer-term evolutionary relationships between strains [38] and so could help to resolve the lineage of these non-G1407 isolates.

Treatment failure in patients infected with gonococci exhibiting cefixime MICs of the levels in the current study have been documented in the UK [10], Norway [9] and Austria [6] and all belonged to G1407. One ST1407 isolate with high-level resistance to cefixime (MIC=4 mg/L) and ceftriaxone (MIC=1–2 mg/L), also associated with treatment failure using cefixime, has been described from France [11]. This strain demonstrates that the ST1407 clone can by a single additional mutation in its penicillin-binding protein 2 (*penA*) mosaic gene develop high-level resistance to both cefixime and ceftriaxone, which is the last remaining option for single antimicrobial empiric treatment of gonorrhoea [11]. The observed predominance of ST1407 and related STs in many EU/EEA countries is therefore worrying given their potential to be therapeutically challenging. While shown to be associated with MSM, it is evident that G1407 is circulating in the heterosexual population also and, accordingly, the risk of future treatment failure is not restricted to any one patient group. G1407 therefore represents a potential major public health problem if it continues to disseminate without any measures taken to restrict this.

G225 was the most prevalent genogroup after G1407 in the current study. ST225 has been reported as a highly prevalent strain in EU/EEA countries [22,25,33,34], with a confirmed association with ciprofloxacin resistance [22,25,33]. Given that most studies began sampling as early as 2003 to 2005, the continued persistence of G225 by 2010 indicates that it is a highly successful, stable NG-MAST clone. Interestingly, a recent Canadian study of isolates showing decreased susceptibility to cefixime and/or ceftriaxone showed that 19% of isolates were ST225 [21]. All isolates of G225 in the current study were sensitive to cefixime and ceftriaxone, but the potential for this strain to show decreased susceptibility is of concern given its potential to predominate and persist. While strong associations between ST225 and MSM are reported [25,34], these were not confirmed in all studies [39] including this one. This could indicate a shift in the distribution of G225 since the earlier typing studies, with possible bridging between homosexual and heterosexual networks. It is acknowledged however that the completeness of the sexual orientation variable in the current study was low, which may have introduced a bias into this analysis.

None of the other major genogroups defined in the current study showed an association with resistance to the antimicrobials tested, but G2, G25, G359 and G387 would appear to be particularly persistent strains also, all having been reported in a molecular survey in 2004 [34]. Individual STs from these groups have been reported in other studies which confirmed that these were susceptible to the antimicrobials tested [22,25,33]. As observed previously [34], genogroups G2, G25 and G387 were associated with heterosexual patients, demonstrating their persistence within this patient population. In contrast other genogroups (G2992, G1407) were shown to be associated with MSM. While ST1407 has been reported in a high proportion of MSM previously [13], to our knowledge this is the first time ST2992 and related types has been reported as a potentially MSM associated type. None of the molecular epidemiological surveys conducted to date have reported ST2992 as a prevalent ST [24-26,28,33,34]. G2992 accounted for 8% of all isolates in our study and while widely distributed, it was particularly predominant in Ireland and Norway, which may indicate an outbreak within MSM, although further longitudinal surveillance of a sufficiently representative sample would be required to verify this.

In our study the selection criteria for isolates were based on accepting from each country consecutive isolates with a full antimicrobial profile. There were no exclusion criteria based on epidemiological characteristics and all patient ages were included. One of the age groups used for the analyses spanned 0 to 24 year-olds however only three of the 363 patients in this group were less than 15 years-old. The inclusion of isolates from the latter patients was valid for the antimicrobial resistance analyses. Only one of the three patients was related to a common genogroup (G225) therefore including these three patients should have had little effect on the overall statistical analyses.

In conclusion, this study shows clear associations between antimicrobial resistance and molecular type. The use of molecular epidemiological typing to predict antimicrobial resistance and follow the spread of antimicrobial resistant gonococcal clones has added public health benefits as it can aid understanding of the dissemination of resistance within a population and facilitate development of targeted intervention strategies, for example by ensuring patients infected with strains likely to be therapeutically challenging are managed with aggressive therapy and test of cure. Additionally if the prediction of antimicrobial resistance was sufficiently reliable this approach could directly impact on appropriate management of patients for whom culture and associated antimicrobial susceptibility testing has not been performed. While it is acknowledged that sample representativeness is a limitation of the current study [12], additional molecular epidemiological typing has successfully demonstrated a 'proof of principle' that this approach could be valid and have real public health value. However for this to be effective, ongoing, frequent molecular typing surveillance will be essential, as there is continual potential for novel STs and novel antimicrobial susceptibility profiles to emerge. Future work should focus on longitudinal typing of a representative sample to monitor stability of associations between ST and antimicrobial resistance and/or epidemiological characteristics, and to identify temporal changes and emergence of novel STs.

Funding

This work was supported by the European Centre for Disease Prevention and Control (call for tender number OJ/2009/03/11-PROC/2009/006), which coordinates the European STI Expert Network, including the STI microbiology programme.

Acknowledgements

We are indebted to all national nominated contact points within the European STI Expert Network and to their colleagues for their continued support and provision of gonococcal isolates as part of the European Gonococcal Antimicrobial Surveillance Programme (Euro-GASP). We thank all members of the European Expert Group on Molecular Typing for their contribution to the original strategy for European typing. We are grateful to Steen Hoffman, Mrs Lene Berthelsen and Miss Ronza Hadad for their invaluable contribution to the antimicrobial susceptibility testing, and to Miss Katy Town for her valued assistance with the statistical analyses.

References

1. European Centre for Disease Prevention and Control (ECDC). Sexually Transmitted Infections in Europe, 1990-2010. Stockholm: ECDC; 2012. Available from: http://www.ecdc.europa.eu/en/publications/Publications/201206-Sexually-Transmitted-Infections-Europe-2010.pdf
2. World Health Organization (WHO). Emergence of multi-drug resistant Neisseria gonorrhoeae - Threat of global rise in untreatable sexually transmitted infections. WHO/RHR/11.14. 2011. Fact sheet. Geneva:WHO; 2011. Available from: http://www.who.int/reproductivehealth/publications/rtis/who_rhr_11_14/en/index.html
3. Lewis DA. The Gonococcus fights back: is this time a knock out? Sex Transm Infect. 2010;86(6):415-21.
4. Bignell CJ; IUSTI/WHO. 2009 European (IUSTI/WHO) guideline on the diagnosis and treatment of gonorrhoea in adults. Int.J.STD AIDS. 2009;20(7), 453-7.
5. Bignell C, Unemo M. 2012 European Guideline on the Diagnosis and Treatment of Gonorrhoea in Adults. The International Union against Sexually Transmitted Infections (IUSTI); 2012:1-24. Available from: http://www.iusti.org/regions/Europe/pdf/2012/Gonorrhoea_2012.pdf
6. Unemo M, Golparian D, Stary A, Eigentler A. First Neisseria gonorrhoeae strain with resistance to cefixime causing gonorrhoea treatment failure in Austria, 2011. Euro Surveill. 2011;16(43):pii=19998. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19998
7. Unemo M, Golparian D, Potočnik M, Jeverica S. Treatment failure of pharyngeal gonorrhoea with internationally recommended first-line ceftriaxone verified in Slovenia, September 2011. Euro Surveill. 2012;17(25):pii=20200. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20200
8. Unemo M, Golparian D, Hestner A. Ceftriaxone treatment failure of pharyngeal gonorrhoea verified by international recommendations, Sweden, July 2010. Euro Surveill. 2011;16(6):pii=19792. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19792
9. Unemo M, Golparian D, Syversen G, Vestrheim DF, Moi H. Two cases of verified clinical failures using internationally recommended first-line cefixime for gonorrhoea treatment, Norway, 2010. Euro Surveill. 2010;15(47):pii=19721. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19721
10. Ison CA, Hussey J, Sankar KN, Evans J, Alexander S. Gonorrhoea treatment failures to cefixime and azithromycin in England, 2010. Euro Surveill. 2011;16(14):pii=19833. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19833
11. Unemo M, Golparian D, Nicholas R, Ohnishi M, Gallay A, Sednaoui P. High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure. Antimicrob Agents Chemother. 2012;56(3):1273-1280.
12. European Centre for Disease Prevention and Control (ECDC). Gonococcal antimicrobial susceptibility surveillance in Europe, 2010. Stockholm:ECDC; 2012. Available from: http://www.ecdc.europa.eu/en/publications/Publications/1206-Gonococcal-AMR.pdf
13. Chisholm SA, Alexander S, Desouza-Thomas L, Maclure-Webster E, Anderson J, Nichols T et al. Emergence of a Neisseria gonorrhoeae clone showing decreased susceptibility to cefixime in England and Wales. J Antimicrob Chemother. 2011;66(11):2509-12.
14. Golparian D, Hellmark B, Fredlund H, Unemo M. Emergence, spread and characteristics of Neisseria gonorrhoeae isolates with in vitro decreased susceptibility and resistance to extended-spectrum cephalosporins in Sweden. Sex Transm Infect. 2010;86(6):454-460.
15. Liao M, Gu WM, Yang Y, Dillon JA. Analysis of mutations in multiple loci of Neisseria gonorrhoeae isolates reveals effects of PIB, PBP2 and MtrR on reduced susceptibility to ceftriaxone. J Antimicrob Chemother. 2011;66(5):1016-23.
16. Lee SG, Lee H, Jeong SH, Yong D, Chung GT, Lee YS et al. Various penA mutations together with mtrR, porB and ponA mutations in Neisseria gonorrhoeae isolates with reduced susceptibility to cefixime or ceftriaxone. J Antimicrob Chemother. 2010;65(4):669-75.
17. Lo JY, Ho KM, Lo AC. Surveillance of gonococcal antimicrobial susceptibility resulting in early detection of emerging resistance. J Antimicrob Chemother. 2012;67(6):1422-6.
18. Tapsall JW, Ray S, Limnios A. Characteristics and population dynamics of mosaic penA allele-containing Neisseria gonorrhoeae isolates collected in Sydney, Australia, in 2007-2008. Antimicrob Agents Chemother. 2010;54(1):554-6.
19. Carannante A, Prignano G, Cusini M, Matteelli A, Dal Conte I, Ghisetti V et al. Cefixime and ceftriaxone susceptibility of Neisseria gonorrhoeae in Italy from 2006 to 2010. Clin Microbiol Infect. 2012;18(6):558-64.
20. Pandori M, Barry PM, Wu A, Ren A, Whittington WL, Liska S et al. Mosaic penicillin-binding protein 2 in Neisseria gonorrhoeae isolates collected in 2008 in San Francisco, California. Antimicrob Agents Chemother. 2009;53(9):4032-4.
21. Martin I, Sawatzky P, Allen V, Hoang L, Lefebvre B, Mina N et al. Emergence and characterization of Neisseria gonorrhoeae isolates with decreased susceptibilities to ceftriaxone and cefixime in Canada: 2001-2010. Sex Transm Dis. 2012;39(4):316-23.
22. Palmer HM, Young H, Graham C, Dave J. Prediction of antibiotic resistance using Neisseria gonorrhoeae multi-antigen sequence typing. Sex Transm Infect. 2008;84(4):280-4.
23. Hjelmevoll SO, Golparian D, Dedi L, Skutlaberg DH, Haarr E, Christensen A et al. Phenotypic and genotypic properties of Neisseria gonorrhoeae isolates in Norway in 2009: antimicrobial resistance warrants an immediate change in national management guidelines. Eur J Clin Microbiol Infect Dis. 2012;31(6):1181-6.
24. Starnino S, Suligoi B, Regine V, Bilek N, Stefanelli P; Neisseria gonorrhoeae Italian Study Group, et al. Phenotypic and genotypic characterization of Neisseria gonorrhoeae in parts of Italy: detection of a multiresistant cluster circulating in a heterosexual network. Clin Microbiol Infect. 2008;14(10):949-54.
25. Florindo C, Pereira R, Boura M, Nunes B, Paulino A, Gomes JP et al. Genotypes and antimicrobial-resistant phenotypes of Neisseria gonorrhoeae in Portugal (2004-2009). Sex Transm Infect. 2010;86(6):449-53.
26. Unemo M, Vorobieva V, Firsova N, Ababkova T, Leniv I, Haldorsen BC et al. Neisseria gonorrhoeae population in Arkhangelsk, Russia: phenotypic and genotypic heterogeneity. Clin Microbiol Infect. 2007;13(9):873-8.
27. Tanaka M, Koga Y, Nakayama H, Kanayama A, Kobayashi I, Saika T et al. Antibiotic-resistant phenotypes and genotypes of Neisseria gonorrhoeae isolates in Japan: identification of strain clusters with multidrug-resistant phenotypes. Sex Transm Dis. 2011;38(9):871-5.
28. Wong WW, Huang CT, Li LH, Chiang CC, Chen BD, Li SY. Molecular epidemiological identification of Neisseria gonorrhoeae clonal clusters with distinct susceptibility profiles associated with specific groups at high risk of contracting human immunodeficiency virus and syphilis. J Clin Microbiol. 2008;46(12):3931-4.
29. Chisholm SA, Quaye N, Cole MJ, Fredlund H, Hoffmann S, Jensen JS et al. An evaluation of gentamicin susceptibility of Neisseria gonorrhoeae isolates in Europe. J Antimicrob Chemother. 2011;66(3):592-5.
30. 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.

31. Buono S, Wu A, Hess DC, Carlson JS, Rauch L, Philip SS et al. Using the Neisseria gonorrhoeae multiantigen sequence-typing method to assess strain diversity and antibiotic resistance in San Francisco, California. Microb Drug Resist. 2012;18(5):510-7.
32. Hess D, Wu A, Golparian D, Esmaili S, Pandori W, Sena E et al. Genome sequencing of a Neisseria gonorrhoeae isolate of a successful international clone with decreased susceptibility and resistance to extended-spectrum cephalosporins. Antimicrob Agents Chemother. 2012;56(11):5633-41.
33. Olsen B, Hadad R, Fredlund H, Unemo M. The Neisseria gonorrhoeae population in Sweden during 2005-phenotypes, genotypes and antibiotic resistance. APMIS. 2008;116(3):181-9.
34. Risley CL, Ward H, Choudhury B, Bishop CJ, Fenton KA, Spratt BG et al. Geographical and demographic clustering of gonorrhoea in London. Sex Transm Infect. 2007;83(6):481-7.
35. Eastick K. Gonococcal antibiotic surveillance in Scotland (GASS): prevalence, patterns and trends in 2009. Health Protection Scotland Weekly Report. 2010;44(34).
36. Huang CT, Yen MY, Wong WW, Li LH, Lin KY, Liao MH et al. Characteristics and dissemination of mosaic penicillin-binding protein 2-harboring multidrug-resistant Neisseria gonorrhoeae isolates with reduced cephalosporin susceptibility in northern Taiwan. Antimicrob Agents Chemother. 2010;54(11):4893-5.
37. Ohnishi M, Golparian D, Shimuta K, Saika T, Hoshina S, Iwasaku K et al. Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. Antimicrob Agents Chemother. 2011;55(7):3538-45.
38. Vidovic S, Horsman GB, Liao M, Dillon JA. Influence of conserved and hypervariable genetic markers on genotyping circulating strains of Neisseria gonorrhoeae. PLoS One. 2011;6(12):e28259.
39. Martin IM, Ison CA, Aanensen DM, Fenton KA, Spratt BG. Changing epidemiologic profile of quinolone-resistant Neisseria gonorrhoeae in London. J Infect Dis. 2005;192(7):1191-5.

A snapshot of genetic lineages of *Mycobacterium tuberculosis* in Ireland over a two-year period, 2010 and 2011

M M Fitzgibbon (MFitzgibbon@STJAMES.IE)[1], N Gibbons[1], E Roycroft[1,2], S Jackson[3], J O'Donnell[3], D O'Flanagan[3], T R Rogers[1,2]
1. Irish Mycobacteria Reference Laboratory, St. James' Hospital, Dublin, Ireland
2. Department of Clinical Microbiology, Trinity College, Dublin, Ireland
3. Health Protection Surveillance Centre, Dublin, Ireland

Citation style for this article:
Fitzgibbon MM, Gibbons N, Roycroft E, Jackson S, O'Donnell J, O'Flanagan D, Rogers TR. A snapshot of genetic lineages of Mycobacterium tuberculosis in Ireland over a two-year period, 2010 and 2011. Euro Surveill. 2013;18(3):pii=20367. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20367

Article submitted on 09 July 2012 / published on 17 January 2013

Mycobacterial interspersed repetitive-unit-variable-number tandem repeat typing alone was used to investigate the genetic lineages among 361 *Mycobacterium tuberculosis* strains circulating in Ireland over a two-year period, 2010 and 2011. The majority of isolates, 63% (229/361), belonged to lineage 4 (Euro-American), while lineages 1 (Indo-Oceanic), 2 (East-Asian) and 3 (East-African–Indian) represented 12% of isolates each (42/361, 45/361, and 45/361, respectively). Sub-lineages Beijing (lineage 2), East-African–Indian (lineage 1) and Delhi/central-Asian (lineage 3) predominated among foreign-born cases, while a higher proportion of Euro-American lineages were identified among cases born in Ireland. Eighteen molecular clusters involving 63 tuberculosis (TB) cases were identified across four sub-lineages of lineage 4. While the mean cluster size was 3.5 TB cases, the largest cluster (involving 12 Irish-born cases) was identified in the Latin American–Mediterranean sub-lineage. Clustering of isolates was higher among Irish-born TB cases (47 of 63 clustered cases), whereas only one cluster (3/63) involved solely foreign-born individuals. Four multidrug-resistant cases identified during this period represented lineages 2 and 4. This study provides the first insight into the structure of the *M. tuberculosis* population in Ireland.

Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* remains a serious challenge to public health worldwide. Despite an overall decline in case notification rates for TB across Europe, rates vary significantly, with the highest rates reported from eastern Europe and the Baltic States [1]. Multidrug-resistant (MDR) TB continues to be a major problem and an added burden in high-incidence countries such as Romania (108.2/100,000) and the Baltic states Lithuania (62.1/100,000), Latvia (43.2/100,000) and Estonia (30.7/100,000) [1]. Ireland has a low incidence, with notification rates that ranged between 9.7 and 11.3 per 100,000 population between 2001 and 2010 although to some extent this may have been influenced by migrants arriving from high-burden countries [1-2]. In Ireland, TB is a statutorily notifiable disease, and in a recent report on the epidemiology of TB in the country, the proportion of culture-confirmed TB cases was 71.2% in 2009 and 63.2% in 2010 (data from 2010 were not finalised at the time of submission) [3]. These proportions are similar to those reported for previous years [3].

Disruption of transmission chains is a key factor in controlling TB both at a national and international level [4]. In recent years, there have been significant advances in developing the molecular tools required for rapid diagnosis of TB [4]. Analysis of variable-number tandem repeat (VNTR) sequences at mycobacterial interspersed repetitive units (MIRU) has emerged as a valuable marker for genotyping strains of the *M. tuberculosis* complex [5]. In large population-based studies, MIRU-VNTR typing has been shown to have similar discriminatory power when compared to IS*6110* restriction fragment length polymorphism (RFLP) typing [5-7]. An optimised set of 24 MIRU-VNTR markers has become the gold standard for genotyping *M. tuberculosis* complex strains worldwide [5,7]. MIRU-VNTR typing is a PCR-based method that yields rapid, reproducible results that are expressed as a 24-digit numerical code which allows for easy exchange of data [5-8]. This method can be applied to early mycobacterial cultures and more recently has been successfully applied directly to smear-positive specimens [5,9]. Previous studies have shown MIRU-VNTR typing to be useful in comparing strains (i) at national and international level, (ii) among household contacts, (iii) associated with drug resistance and (iv) to determine the evolutionary pathway of TB [5-6,10-16]. At European level, MIRU-VNTR typing has been adopted for the molecular surveillance of the international transmission of MDR-TB and extensively drug-resistant TB [7].

In November 2009, molecular genotyping in the form of 24-locus MIRU-VNTR typing was introduced at the

Irish Mycobacteria Reference Laboratory (IMRL) which both cultures and receives isolates from microbiology laboratories around the country. To allow for rapid, high-throughput genotyping of *M. tuberculosis*, the commercial MIRU-VNTR typing kit (GenoScreen, Lille, France) was introduced in 2010 [5,7]. As 24-locus MIRU-VNTR typing is considered the gold standard genotyping method, all *M. tuberculosis* isolates identified at the IMRL are currently typed prospectively with this method, and it is envisaged that all *M. tuberculosis* isolates recovered since 2000 will be typed on a retrospective basis.

Here we report the first analysis of the structure of the *M. tuberculosis* population in Ireland for isolates recovered during 2010 and 2011 following the introduction of 24-locus MIRU-VNTR typing to the diagnostic laboratory. It needs to be noted that at the time there was an under-representation of isolates from the southern region of Ireland.

Methods

MIRU-VNTR typing

M. tuberculosis isolates (n=361) recovered in or referred to the IMRL over a two-year period (2010–11) were typed with the MIRU-VNTR typing kit (GenoScreen) [5]. Validation of the MIRU-VNTR technique was performed using the MIRU-VNTR Calibration Kit (GenoScreen). PCR products were subjected to electrophoresis using a 3130 genetic analyser (Applied Biosystems). Sizing of fragments and MIRU-VNTR allele assignation was performed using GeneMapper software (Applied Biosystems). Phylogenetic lineages were assigned to each isolate using the MIRU-VNTR*plus* online tool [17-18].

The 24-locus MIRU-VNTR panel comprised the following loci: MIRU 02, VNTR 42, VNTR 43, MIRU 04, MIRU 40, MIRU 10, MIRU 16, VNTR 1955, MIRU 20, QUB 11b, ETR A, VNTR 46, VNTR 47, VNTR 48, MIRU 23, MIRU 24, MIRU 26, MIRU 27, VNTR 49, MIRU 31, VNTR 52, QUB 26, VNTR 53, and MIRU 39. The MIRU-VNTR profiles are reported as a series of 24 numbers that correspond to the number of alleles at each of the loci described above.

Clusters of isolates were defined as two or more isolates with indistinguishable MIRU-VNTR patterns. The strain clustering rate was calculated as $(n_c\text{-}c)/n$, where n_c was the total number of strain-clustered cases, c was the number of clusters and n was the total number of isolates [5].

Epidemiological analysis

Enhanced surveillance of TB was implemented in Ireland in 1998. Enhanced TB notification forms are completed by public health doctors, summarising all available clinical, microbiological, histological and epidemiological information. These data are collated in the regional public health departments. Anonymised data are then submitted electronically to the Health Protection Surveillance Centre (HPSC) for the production of reports on a weekly, quarterly and annual basis.

Since January 2011, cases of TB have been reported through the Computerised Infectious Disease Reporting system (CIDR). CIDR is a web-based system developed to integrate case-based clinical and laboratory data in order to manage the surveillance and control of notifiable infectious diseases in Ireland. Prior to using CIDR for TB surveillance, MIRU-VNTR typing results were not linked to case-based epidemiological data. In addition to recording sporadic case-based data, CIDR also facilitates the reporting of clustered cases, according to Irish outbreak case definitions [2]. Clustered cases can be reported via a summary aggregate outbreak data module to which the relevant disaggregate case-based surveillance data can also be linked.

Data analysis

The TB enhanced surveillance data for 2011 (epidemiological and linked laboratory data) used in this publication were extracted from CIDR on 17 April 2012 using Business Objects XI software and were analysed using Microsoft Excel. Data for 2011 were provisional at the time of extraction and subject to ongoing validation and revision.

Results

Results are presented in two separate sections. In the first part, genotyping results for isolates recovered in 2010–11 are presented. As epidemiological data linking was available from 2011 onwards, the second part of the results section (enhanced surveillance) refers to genotyping results linked to epidemiological data for 2011 isolates only.

Mycobacterium tuberculosis genotyping, 2010–11

Some 361 *M. tuberculosis* isolates were recovered in or referred to the IMRL during 2010–11, representing 63.6% of culture-positive cases identified through the national TB surveillance system in that period. Genotyping of *M. tuberculosis* isolates recovered during the study period yielded four global lineages (Table 1). The majority (63%) belonged to lineage 4 (Euro-American), while lineages 1 (Indo-Oceanic), 2 (East-Asian) and 3 (East-African–Indian) represented 12% each. Among the 229 Euro-American strains, sub-lineages Latin American–Mediterranean (LAM) (23%), Haarlem (21%), H37Rv (19%) and Haarlem/X (13%) were most prevalent (Table 1).

Within lineage 4, 18 clusters were identified involving 63 TB cases (Table 2). The strain clustering rate varied between different sub-lineages, but was highest for the LAM sub-lineage (6.9%). While the mean cluster size was 3.5 TB cases, the largest cluster (involving 12 Irish-born cases, representing 19% of all clustered isolates) was identified within the LAM sub-lineage (Table

Table 1

Distribution of lineages among *Mycobacterium tuberculosis* isolates, Ireland, 2010–11 (n=361)

Global lineage	Sub-lineage	No. of isolates	% of isolates
1 Indo-Oceanic	East-African–Indian	42	12
2 East-Asian	Beijing	45	12
3 East–African–Indian	Delhi/central-Asian	45	12
4 Euro-American	Lineage 4 total	229	63
	Latin American–Mediterranean	52	
	Haarlem	47	
	H37Rv	44	
	Haarlem/X	29	
	Cameroon	13	
	S	6	
	TUR	8	
	X	5	
	Ghana	3	
	URAL	3	
	Uganda I & II	6	
	NEW-1	2	
	Others[a]	11	
Total		**361**	**100**

[a] The category Others includes isolates for which the sub-lineages were not clearly defined

2). Four other clusters within the LAM sub-lineage contained between 4.7% (3/63) and 12.7% (8/63) of clustered cases. Among the clustered cases of Haarlem, H37Rv and Haarlem/X, cluster sizes ranged from 3.2 to 9.5%, at 3.2% and from 3.2 to 4.7% of isolates, respectively.

Only one cluster (3/63) contained exclusively foreign-born individuals, 12 clusters (47/63) involved Irish-born cases only, while five clusters (13/63) were mixed. In addition, one small cluster was observed among the isolates from lineage 2.

The four MDR-TB cases identified during this period represented lineages 2 (Beijing) and 4 (Ural, H37Rv and LAM). None of the MDR-TB cases were clustered.

Tuberculosis enhanced surveillance data for 2011 isolates (epidemiological and laboratory)

In 2011, 432 TB cases were provisionally reported on CIDR, of which approximately 166 (38%) were typed. At the time of data extraction, 136 TB cases were updated to include MIRU typing results (representing 81.9% of 166 typed isolates). Of the 136 TB cases with a MIRU typing result, 34 were clustered in 11 clusters with different MIRU types. Clusters ranged in size from eight to two TB cases. Of the 11 MIRU type clusters, five, comprising 18 TB cases, were confirmed by public health departments as outbreaks meeting the Irish case definition.

The Beijing sub-lineage was most prevalent (15.4%) and associated with a small cluster. Sub-lineages Haarlem, LAM, and H37Rv were most prevalent among lineage 4 strains, while lineage 1 and lineage 3 represented 11.8% and 10.3% of typed isolates, respectively. Interestingly, isolates recovered from pulmonary specimens were mostly correlated with lineage 4 strains, while the majority of isolates recovered from extra-pulmonary specimens belonged to lineages 1 and 3 (Figure 1). In lineage 3, nine of 14 isolates were recovered from patients born in Pakistan, while the remaining five isolates were recovered from patients born in India (n=2), Kenya (n=1), Nepal (n=1) and Nigeria (n=1). Only one lineage 1 isolate was recovered from an Irish-born patient, while six were recovered from patients born in the Philippines. Other countries represented among lineage 1 isolates were Bangladesh, India, Mozambique, Pakistan, Somalia and Vietnam.

The distribution of lineages among Irish-born and foreign-born TB cases is shown in Figure 2. Of the 127 TB cases for whom MIRU-VNTR and country of birth were known, 51.5% were foreign-born and 41% were Irish-born. Lineages 1, 2 and 3 predominated among foreign-born TB cases, while a higher proportion of lineage 4 isolates were identified among Irish-born cases.

Discussion

This study has provided a snapshot of the genetic diversity of *M. tuberculosis* in Ireland. Due to the small numbers of isolates in our study, statistical analysis would not be significant and was not performed. Although data on sub-lineages were analysed by age and sex, the resulting frequencies were too small to draw firm conclusions from. However, when age and sex analyses were further stratified by country of birth, these data were broadly similar to the age and sex profile of the Irish TB notification data.

A large diverse group of isolates has been identified, suggesting a low degree of active transmission among TB patients. The distribution of genetic lineages is similar to other recent studies that used different typing techniques and in which lineage 4 (Euro-American) predominated among circulating *M. tuberculosis* strains [12,19-21]. In previous work conducted in the south-west region of Ireland, lineage 4 predominated, and clustering of isolates was associated with Irish nationals and lineage 4 isolates only [22]. In our study, the distribution of genetic lineages among extra-pulmonary specimens (where lineages 1 and 3 predominated) was similar to a recently published large-scale study conducted in the United States (US) investigating the relationship between genetic lineages and clinical sites of infection [23]. In the US study, the highest percentage of isolates recovered from extra-pulmonary

Table 2

Clusters of *Mycobacterium tuberculosis* isolates within lineage 4 (Euro-American), Ireland, 2010–11 (n=172 isolates)

Sub-lineage	Total no. of cases	No. of clustered cases (%)	No. of clusters	Strain clustering rate (%)	No. of isolates/ cluster (% clustered cases)	MIRU-VNTR profile[a]
Latin American–Mediterranean	52	30 (8.3)	5	6.9	3 (4.7)	124244332224126153332832
					8 (12.7)	142244332224126153322622
					12 (19)	142244332224126143322622
					3 (4.7)	132244332224125153322222
					4 (6.3)	132244332224126133322622
Haarlem	47	18 (5)	6	3.3	2 (3.2)	223225342334425143323332
					6 (9.5)	22322534233442515332 3_32
					2 (3.2)	123235332634425143423332
					4 (6.3)	22322534233442514332 3_32
					2 (3.2)	223235331532423153333632
					2 (3.2)	223215332334425153323632
H37Rv	44	8 (2.2)	4	1.1	2 (3.2)	224243122234225153234422
					2 (3.2)	22424312243422515333 5512
					2 (3.2)	224213222534226153334522
					2 (3.2)	224213322334226153335522
Haarlem/X	29	7 (1.9)	3	1.1	2 (3.2)	224234342334425154135832
					3 (4.7)	243244332434425153343a32
					2 (3.2)	243234332234425143331832
Total clustered lineage 4 cases	**172**	**63**	**18**	-	-	-

[a] The numbers in this 24-digit profile correspond to the number of alleles at each of the following loci: MIRU 02, VNTR 42, VNTR 43, MIRU 04, MIRU 40, MIRU 10, MIRU 16, VNTR 1955, MIRU 20, QUB 11b, ETR A, VNTR 46, VNTR 47, VNTR 48, MIRU 23, MIRU 24, MIRU 26, MIRU 27, VNTR 49, MIRU 31, VNTR 52, QUB 26, VNTR 53, MIRU 39.

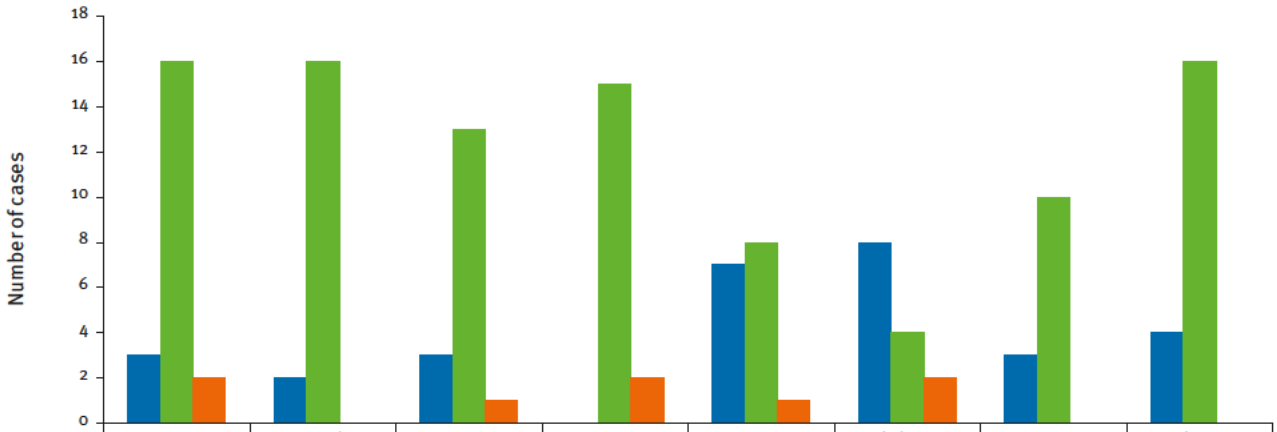
specimens was from lineages 1 (22.6%) and 3 (34.3%) [23]. However, due to the small numbers of exclusive extra-pulmonary specimens (n=30) and limited epidemiological data, statistical analysis of the relationship between lineage and clinical site of infection was not possible in our report.

Molecular clustering of isolates in our study was more common among Irish-born individuals. These findings were similar to a previous Irish study conducted by Ojo et al. in the south-west region of Ireland, but unlike a recent study conducted in Switzerland [12,22]. We identified 18 clusters in lineage 4, and the mean cluster size was 3.5 TB cases. The largest cluster, involving 12 TB cases, belonged to the LAM lineage and spanned a period of 18 months. A second cluster identified in the LAM lineage differed by a single locus variant (SLV) at locus 2996. Similarly, in the Haarlem lineage, the two largest clusters differed by a SLV at locus 2996 also. MIRU 26 (or locus 2996) has yielded stable comparable results in a large-scale study investigating 824 *M. tuberculosis* isolates conducted at the Institut Pasteur de Lille, France, in 2006 [5]. Molecular typing played a key role in identifying a dominant *M. tuberculosis* strain (known as the Mercian strain) circulating in the West Midlands region in the United Kingdom (UK) over a five-year period, highlighting the importance of cluster analysis [14]. Prospective molecular typing can identify rapidly expanding clusters of *M. tuberculosis* before they spread further into the community. A single dominant MIRU-VNTR type was not observed in this study, however, this could be due to the study period being short. In contrast, prospective molecular typing of *M. tuberculosis* by RFLP, performed since 1993 in the Netherlands, has proven to be effective. DNA fingerprinting data has been shown to be a powerful tool in defining epidemiological links and guiding TB control programmes in the Netherlands [24-25].

Another limitation of this study is that only one typing method was used to investigate the *M. tuberculosis* population structure in Ireland. Previous studies have shown that a combination of MIRU-VNTR typing and spoligotyping can differentiate more readily between *M. tuberculosis* strains [26-27]. However, in a previous Irish study using both spoligotyping and MIRU-VNTR typing, MIRU-VNTR typing identified clusters among spoligotype groups, thus providing supporting evidence that MIRU-VNTR typing is a more discriminatory typing method [22]. The discriminatory power of the 24-locus MIRU-VNTR panel used in this study has shown to be similar to IS*6110* RFLP analysis [5]. However, the discriminatory power of 24-locus MIRU-VNTR typing differs among genetic lineages, and the inclusion of

Figure 1

Distribution of *Mycobacterium tuberculosis* lineages by site of infection, Ireland, 2011 (n=136)

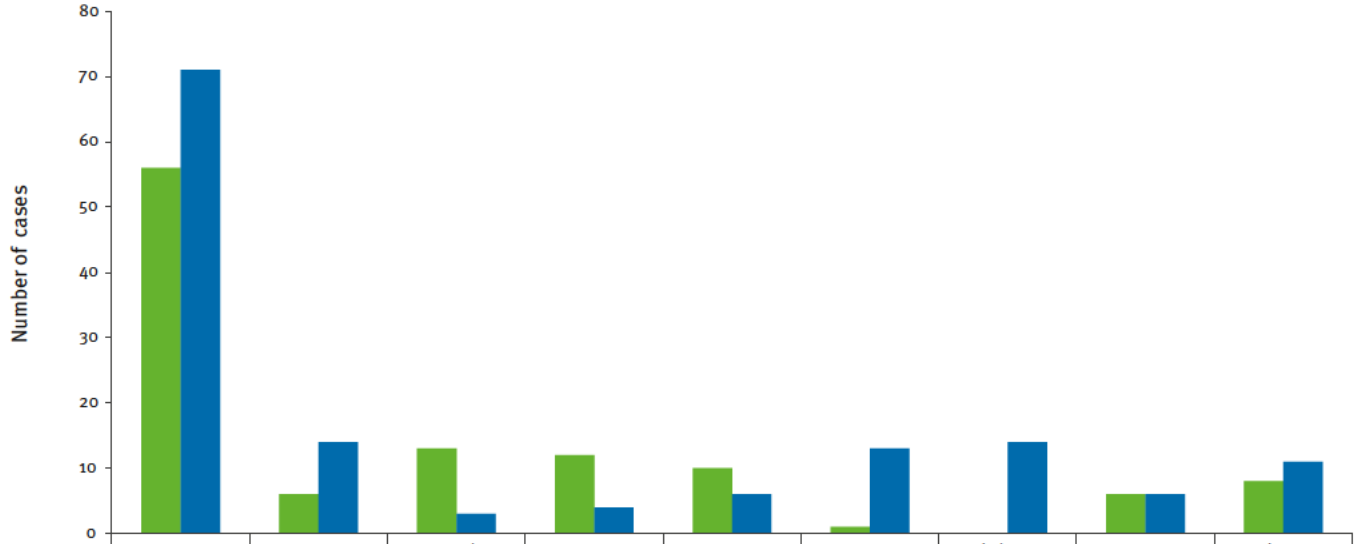


	Beijing n=21	Haarlem/X n=18	Haarlem n=17	LAM n=17	EAI n=16	Delhi/CAS n=14	H37Rv n=13	Others n=20
Extra-pulmonary	3	2	3	0	7	8	3	4
Pulmonary	16	16	13	15	8	4	10	16
Both	2	0	1	2	1	2	0	0

CAS: central-Asian; EAI: east-African–Indian; LAM: Latin American–Mediterranean.

Figure 2

Distribution of lineages among Irish and non-Irish typed *Mycobacterium tuberculosis* cases, Ireland, 2011 (n=127)



	Overall (n=127)	Beijing n=20	Haarlem/X n=16	Haarlem n=16	LAM n=16	EAI n=14	Delhi/CAS n=14	H37Rv n=12	Others n=19
Irish	56	6	13	12	10	1	0	6	8
Non-Irish	71	14	3	4	6	13	14	6	11

CAS: central-Asian; EAI: east-African–Indian; LAM: Latin American–Mediterranean.

additional hypervariable loci may be required to differentiate among strains of lineages 2 (Beijing) and 3 (Delhi/central-Asian). For enhanced cluster or outbreak analysis, whole-genome sequencing has been shown to differentiate among strains with identical 24-locus MIRU-VNTR patterns [28-29]. The role of whole-genome sequencing in investigating community outbreaks in the UK was reported recently [29]. Walker et al. estimated that the rate of genetic changes was 0.5 single nucleotide polymorphisms (SNPs) per genome per year. Furthermore, the maximum number of genetic changes over three years would be five SNPs and 10 SNPs over 10 years [29]. It has also been proposed that clustering of isolates increases over longer periods as transmission chains are more efficiently analysed and reported [30]. But the *M. tuberculosis* genotype involved in the cluster must be considered as for example the Beijing lineage has increased ability to spread and cause disease. While clustering was limited in our study, the study period was too short to draw clear conclusions.

Although the reproducibility of MIRU-VNTR typing has been well documented, results from the first worldwide proficiency study on this method were surprising [7]. Intra- and inter- laboratory reproducibility varied depending on the typing methods employed in each laboratory. In our setting, when the commercial MIRU-VNTR typing kit was used to analyse the quality control panel, 100% concordance was achieved with the reference data (30/30 tested strains) and 100% intra-laboratory reproducibility was achieved. These findings are important to consider when typing data is exchanged between laboratories.

Although six of the 11 MIRU typing clusters identified during 2011 were not confirmed as outbreaks by public health departments, it is possible that the reason why four of these clusters did not meet the Irish TB outbreak case definitions was the small number (n=2) of involved cases [2].

In summary, this study has provided the first insights into the structure of the *M. tuberculosis* population in Ireland. Although the incidence of TB has remained static in Ireland over the last decade, there has been mass immigration to this island nation. Not surprisingly, lineage 4 predominated among circulating strains of *M. tuberculosis* in the present study. But the degree of diversity among *M. tuberculosis* was unexpected. Future studies in the IMRL involving retrospective genotyping analysis of *M. tuberculosis* isolates collected since 2000 may provide an interesting epidemiological picture. Continued molecular surveillance is important as it has been suggested that the transmissibility profile of *M. tuberculosis* strains may be influenced by their genetic and evolutionary background. This understanding of the dynamics of *M. tuberculosis* strains will provide novel insights into the *M. tuberculosis* population structure and how it relates to the epidemiology of TB in Europe and beyond.

Acknowledgments

The authors would like to thank all the departments of public health, clinicians and laboratories for providing the surveillance data on these TB cases.

References

1. European Centre for Disease Prevention and Control (ECDC). Annual Epidemiological Report 2011. Reporting on 2009 surveillance data and 2010 epidemic intelligence data. Stockholm: ECDC; 2011. Available from: http://www.ecdc.europa.eu/en/publications/Publications/Forms/ECDC_DispForm.aspx?ID=767
2. Health Protection Surveillance Centre (HPSC). Guidelines on the Prevention and Control of Tuberculosis in Ireland 2010.. Dublin: HPSC; 2010. Available from: http://www.hpsc.ie/hpsc/AboutHPSC/ScientificCommittees/Publications/File,4349,en.pdf
3. Health Protection Surveillance Centre (HPSC). Report on the Epidemiology of TB in Ireland, 2009. Dublin: HPSC, 2012. [Accessed 17 Jan 2013]. Available from: http://www.ndsc.ie/hpsc/A-Z/VaccinePreventable/TuberculosisTB/Epidemiology/SurveillanceReports/2009/File,13261,en.pdf
4. Barnes PF. Cave MD. Molecular epidemiology of tuberculosis. New Engl J Med. 2003;349(12):1149-56.
5. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol. 2006;44(12):4498-510.
6. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P. One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med. 2011;105 Suppl 1:S67-73.
7. de Beer J, Kremer K, Ködmön C, Supply P, van Soolingen D, Global Network for the Molecular Surveillance of Tuberculosis 2009. First worldwide proficiency study on variable-number tandem-repeat typing of Mycobacterium tuberculosis complex strains. J Clin Microbiol. 2012;50(3):662-9.
8. Kanduma E, McHugh TD, Gillespie SH. Molecular methods for Mycobacterium tuberculosis strain typing: a user's guide. J Appl Microbiol. 2003;94(5):781-91.
9. Alonso M, Herranz M, Martinez Lirola M, Gonzalez-Rivera M, Bouza E, Garcia de Viedma D. Real-time molecular epidemiology of tuberculosis by direct genotyping of smear-positive clinical specimens. J Clin Microbiol. 2012;50(5):1755-7.
10. Augustynowicz-Kopeć E, Jagielski T, Kozińska M, Kremer K, Van Soolingen D, Bielecki J, et al. Transmission of tuberculosis within family-households. J Infect. 2012;64(6):596-608.
11. Prodinger WM, Polanecký V, Kozáková B, Müllerová M, Mezenský L, Kaustová J, et al. Molecular epidemiology of tuberculosis in the Czech Republic, 2004: analysis of M. tuberculosis complex isolates originating from the city of Prague, south Moravia and the Moravian-Silesian region. Cent Eur J Public Health. 2006;14(4):168-74.
12. Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, Pfyffer GE, et al. Mycobacterium tuberculosis transmission in a country with low incidence: role of immigration and HIV infection. J Clin Microbiol. 2012;50(2):388-95.
13. Maguire H, Brailsford S, Carless J, Yates M, Altass L, Yates S, et al. Large outbreak of isoniazid-monoresistant tuberculosis in London, 1995 to 2006: case-control study and recommendations. Euro Surveill. 2011;16(13):pii=19830. Available from: http://www.eurosurveillance.org/Viewarticle.aspx?ArticleId=19830
14. Evans JT, Serafino Wani RL, Anderson L, Gibson AL, Grace Smith E, Wood A, et al. A geographically-restricted but prevalent Mycobacterium tuberculosis strain identified in the West Midlands region of the UK between 1995 and 2008. PLoS One. 2011; 6(3):e17930.
15. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog. 2008;4(9):e1000160.
16. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, Hafeez I, et al. Molecular epidemiology of Mycobacterium tuberculosis in East Lancashire 2001-2009. Thorax. 2011;66(8):709-13.
17. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res. 2010;38:W326-31.
18. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a

multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. J Clin Microbiol. 2008;46(8):2692-9.

19. Roetzer A, Sieglinde S, Diel R, Gasau F, Ubben T, di Nauta A, et al. Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. J Clin Microbiol. 2011;49(12):4173-8.
20. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:23.
21. Homolka S, Post E, Oberhauser B, George AG, Westman L, Dafae D, et al. High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. BMC Microbiol. 2008;8:103.
22. Ojo OO, Sheehan S, Corcoran D, Nikolayevsky V, Brown T, O'Sullivan M, et al. Molecular epidemiology of Mycobacterium tuberculosis isolates in Southwest Ireland. Infect Genet Evol. 2010;10(7):1110-6.
23. Click ES, Moonan PK, Winston CA, Cowan LS, Oeltmann JE. Relationship between Mycobacterium tuberculosis phylogenetic lineage and clinical site of tuberculosis. Clin Infect Dis. 2012;54(2):211-9.
24. Borgdorff MW, Nagelkerke NJD, de Haas PEW, van Soolingen D. Transmission of Mycobacterium tuberculosis depending on the age and sex of source cases. Am J Epidemiol. 2001;154(10):934-43.
25. Lambregts-van Weezenbeek CS, Sebek MM, van Gerven PJ, de Vries G, Verver S, et al. Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. Int J Tuberc Lung Dis. 2003;7(12 Suppl 3):S463-70.
26. Oelemann MC, Diel R, Vatin V, Haas W, Rüsch-Gerdes S, Locht C, et al. Assessment of an optimized mycobacterial interspersed repetitive-unit-variable number of tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol. 2007;45(3):691-7.
27. Valcheva V, Mokrousov I, Narvskaya O, Rastogi N, Markova N. Utility of new 24-locus variable-number tandem-repeat typing for discriminating Mycobacterium tuberculosis clinical isolates collected in Bulgaria. J Clin Microbiol. 2008;46(9):3005-11.
28. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med. 2011;364(8):730-9.
29. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis. 2012. pii:S1473-3099(12)70277-3. doi: 10.1016/S1473-3099(12)70277-3.
30. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, Godfrey-Faussett P, et al. Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis. European concerted action on molecular epidemiology and control of tuberculosis. Int J Tuberc Lung Dis. 1999;3(12):1055-60.

Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories

H G Niesters1, J W Rossen1,2, H van der Avoort3, D Baas3, K Benschop3,4, E C Claas5, A Kroneman3, N van Maarseveen6, S Pas7, W van Pelt3, J C Rahamat-Langendoen1, R Schuurman6, H Vennema3, L Verhoef3, K Wolthers4, M Koopmans (Marion.Koopmans@rivm.nl)3,7

1. Department of Medical Microbiology, Division of Clinical Virology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
2. Laboratory for Medical Microbiology, St Elisabeth Hospital, Tilburg, the Netherlands
3. Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
4. Department of Medical Microbiology, Academic Medical Center, Amsterdam, the Netherlands
5. Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands
6. Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands
7. Department of Virology, Erasmus Medical Centre, Rotterdam, the Netherlands

Citation style for this article:
Niesters HG, Rossen JW, van der Avoort H, Baas D, Benschop K, Claas EC, Kroneman A, van Maarseveen N, Pas S, van Pelt W, Rahamat-Langendoen JC, Schuurman R, Vennema H, Verhoef L, Wolthers K, Koopmans M. Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories . Euro Surveill. 2013;18(4):pii=20387. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20387

Article submitted 28 June 2012 / published on 24 January 2013

Laboratory-based surveillance, one of the pillars of monitoring infectious disease trends, relies on data produced in clinical and/or public health laboratories. Currently, diagnostic laboratories worldwide submit strains or samples to a relatively small number of reference laboratories for characterisation and typing. However, with the introduction of molecular diagnostic methods and sequencing in most of the larger diagnostic and university hospital centres in high-income countries, the distinction between diagnostic and reference/public health laboratory functions has become less clear-cut. Given these developments, new ways of networking and data sharing are needed. Assuming that clinical and public health laboratories may be able to use the same data for their own purposes when sequence-based testing and typing are used, we explored ways to develop a collaborative approach and a jointly owned database (TYPENED) in the Netherlands. The rationale was that sequence data – whether produced to support clinical care or for surveillance –can be aggregated to meet both needs. Here we describe the development of the TYPENED approach and supporting infrastructure, and the implementation of a pilot laboratory network sharing enterovirus sequences and metadata.

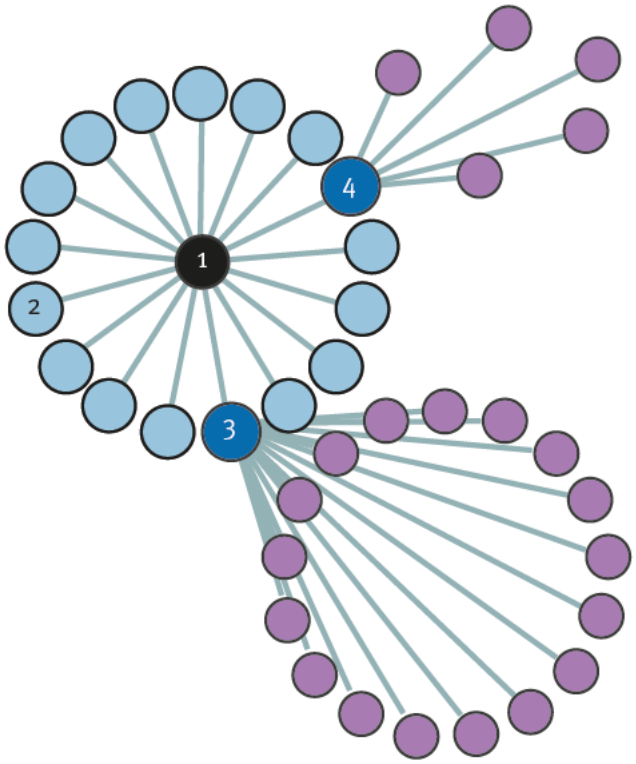
Introduction

Laboratory-based surveillance is one of the pillars of monitoring infectious disease trends, which is based on data from clinical and/or public health laboratories. This type of surveillance is performed for a range of food- and waterborne, sexually transmitted and blood-borne diseases, respiratory pathogens or zoonotic pathogens and provides important input for national and international disease surveillance, to evaluate the impact of control and prevention measures, and to detect clusters or relevant changes in pathogen presence and/or behaviour [1-3].

One problem in the use of laboratory-based surveillance systems is that they require information that typically is collected at the clinical level and therefore is not focused on surveillance. For certain priority diseases, such as polio and measles, this issue has been solved by making the identification of a case notifiable, in which case the laboratory or the clinician or both are required to provide structured information for surveillance to a national or international dedicated organisation. For non-notifiable diseases, however, the need for standardisation to ensure data comparability between laboratories may be at odds with the rapid developments in clinical microbiology laboratories [4-6]. In the Netherlands, currently, diagnostic laboratories routinely submit strains or samples to reference laboratories for characterisation and typing. However, with the introduction of molecular diagnostic methods in most of the larger diagnostic centres, the distinction between diagnostic and reference laboratory functions has become less clear-cut. Multiplex real-time PCR and sequence-based detection and typing techniques may be used for clinical diagnosis, to guide treatment (by, for example, resistance profiling, strain characterisation and typing), for hospital infection control and quality management (for cluster detection). The methods and analytical tools employed for these functions potentially overlap with what is needed for national and international or cross-border surveillance. The expected introduction of next generation

Figure 1

Conceptual model for TYPENED, showing laboratories with different capacities



TYPENED: TYPeer netwerk NEDerland [Typing network Netherlands].

1, 2, 3 and 4 represent a specialist laboratory, dealing with, for example, samples from food, water, the environment and animals.

The laboratory capacities range from routine diagnostic functions, diagnostics and typing functions, expert-level services (includes research), and national reference-level functions.

The dark circle indicates the hub from which the molecular platform infrastructure is provided (see Figure 2). Based on areas of expertise (indicated by numbers), coordination of the network activities may be delegated from the national focal point to a local laboratory, while maintaining the common infrastructure.

sequencing techniques in routine diagnostic settings within the next five years is likely to further lift the borders between the previously separated activities across disciplines and domains [7].

While international surveillance networks rely on reference laboratories, and each pathogen or pathogen group has its own network and system, often with centralised data collection, the latest developments are a challenge for these networks. As more and more clinical laboratories perform molecular testing methods, the reference laboratories become dependent on data submission by these laboratories, often with little perceived benefit for the submitting laboratories, considering the extra effort required. We anticipate increasing resistance from clinical laboratories to data requests for surveillance purposes because of these competing priorities.

Given these developments, we consider that new ways of networking of data and data sharing are needed. Assuming that clinical and public health laboratories may be able to use the same data for their own purposes when sequence-based testing and typing are used, we explored ways to develop a collaborative approach and a jointly owned database in the Netherlands. Here we describe the development of the approach and supporting infrastructure, and the implementation of a pilot laboratory network sharing enterovirus sequences and metadata.

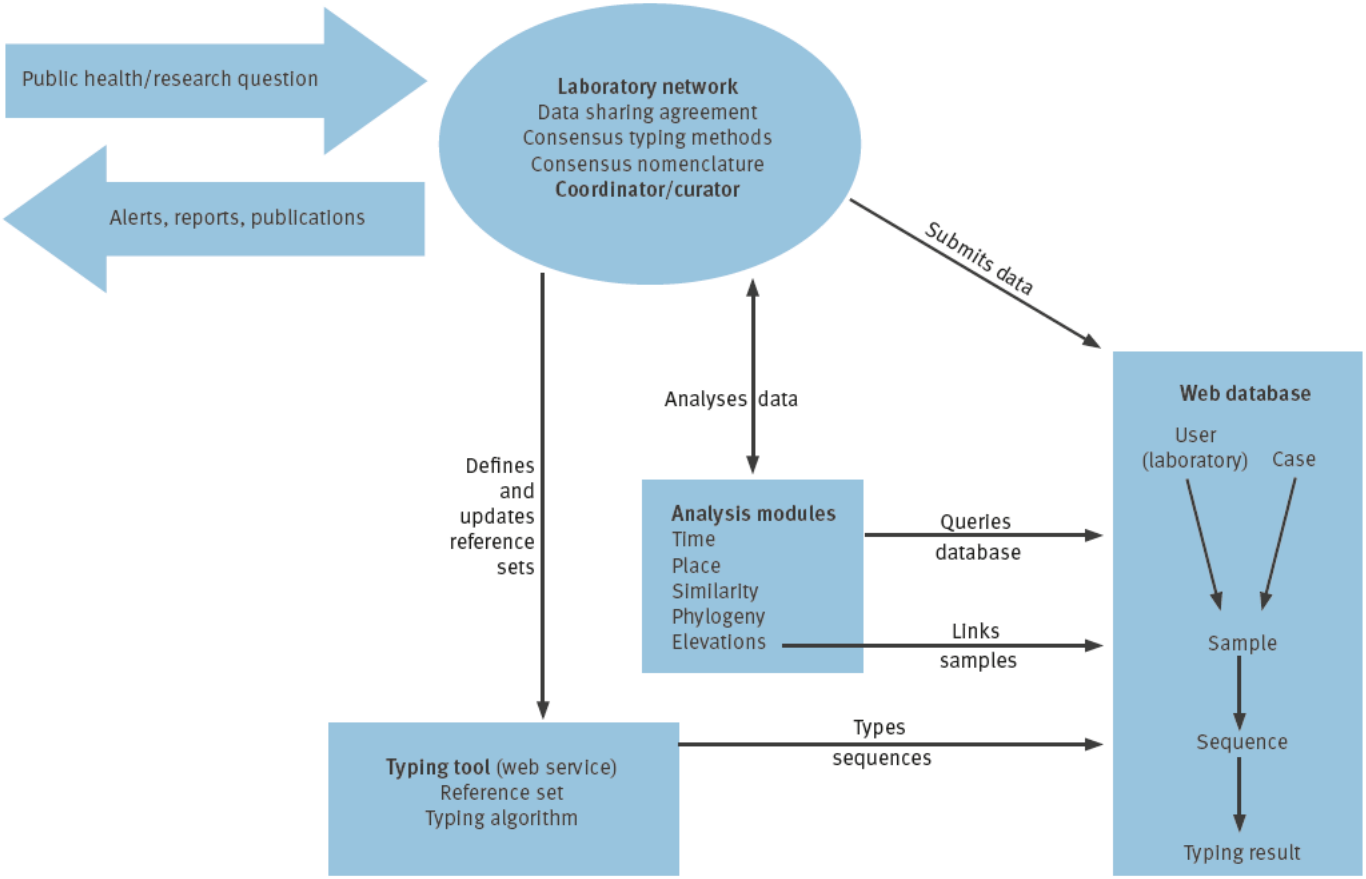
Methods

Partnership

An initiative set up by a group of opinion leaders in microbiology in the Netherlands to draw attention to the changing needs of and demands placed on clinical laboratories and the need for standardisation to ensure data comparability and sharing between laboratories. Within this initiative, called TYPENED (TYPeer netwerk NEDerland [Typing network Netherlands]), two pilots were started in 2009: one for bacterial typing and one for viruses. In the VIRO-TYPENED pilot, five universities and one regional laboratory collaborated with the National Institute of Public Health and the Environment (RIVM) to develop a new model of collaboration for virology based on sequence information gathered in the routine diagnostic setting. These laboratories have all been long-term suppliers of surveillance information, by sending to RIVM isolates or clinical specimens as well as clinical information for a number of viruses such as influenza A virus, norovirus, enterovirus, rotavirus and hepatitis A, B and E viruses. All participating laboratories have molecular diagnostic testing facilities and perform sequencing as part of their routine diagnostics for specific clinical or research questions on one or more of these pathogens (Figure 1, first ring of clinical laboratories surrounding the national reference laboratory). Using a centralised database structure at the national reference laboratory level, expert clinical laboratories can still have

Figure 2

Conceptual model for data sharing platform for TYPENED collaboration between the national public health institute and clinical laboratories in a laboraty network



TYPENED: TYPeer netwerk NEDerland [Typing network Netherlands].

Laboratories submit data to a joint database. The data comprise sequence data and background data on the sample and the patient (case). All sequences are typed automatically. A set of online analysis modules is available for all participants to mine the data. Data can be analysed for trends or clusters in time and place. Sequence data can be analysed for similarity and phylogenetic clustering. Elevations are identified through an automatic cluster detection algorithm based on both sequence information and epidemiological parameters.

their own network activities with collaborating local diagnostic laboratories (Figure 1, showing the group of diagnostic laboratories that refer molecular typing data to the laboratories indicated by '3' or '4').

Selection of pilot pathogens

An inventory was made of the currently used typing methods in the six clinical laboratories and the public health laboratory participating in VIRO-TYPENED using a structured online questionnaire. Participants were asked to list those viruses for which they had typing methods operational in their laboratories and the purpose of those typing applications, and to indicate for which viruses they would like to see joint action, and at which level. The options provided were: (i) the exchange of protocols, control reagents and quality-control panels; (ii) a centralised reference data collection; (iii) a common database; and (iv) no collaboration considered necessary. The purpose of this inventory was to identify areas for which there was a common need, as well as areas where joint action was not considered advantageous. A second part of the inventory asked about methods used and the frequency of typing in each laboratory.

Molecular platform database

In order to achieve efficiency and continuity, a generic database infrastructure for sharing of molecular typing data and metadata was developed at RIVM between 2008 and 2011. The platform consists of a web database and a set of analysis modules. The database can be configured for a specific pathogen, at the request

of a laboratory network, which also appoints a coordinator or curator. User types can be defined, coupled with tailored access rights. The two central entities are sample and sequence. A minimal dataset can be defined by the network, based on the questions addressed, coupled with a feasibility assessment. This dataset minimally comprises time and place, but can be complemented with additional epidemiological or clinical metadata specific to the targeted organism. Besides online data entry forms, the platform provides a bulk upload option using Microsoft Excel and FASTA formats.

All sequences submitted to the database are automatically typed in a standardised way using a web-based typing tool [8]. Sequence data can be analysed by carrying out built-in similarity searches using the BLAST algorithm, and by generating pie charts, incidence plots, geographical maps and phylogenetic trees (neighbour-joining clustering method, with a two-parameter Kimura nucleotide-substitution model, with or without bootstrapping). The added value of a database like this – compared with the database of GenBank [9], in which laboratories all over the world share their sequences – is threefold. Firstly, the data are more comparable because of the agreed typing region and the standardised typing results and secondly, the data are shared before laboratories have decided to make them publicly available, for example, through GenBank. The third important advantage is the linked, standardised set of epidemiological and clinical data with each sequence, which allows in-depth analysis. A description of the components and functions of the molecular platform is shown in Figure 2.

Pilot study: enteroviruses

On the basis of the inventory results, the seven laboratories agreed to start the pilot with enterovirus as a test pathogen. A minimum dataset was agreed, including age and sex of patient, type of sample from which the virus was detected, whether the patient was hospitalised, travel history (by country visited), clinical symptoms in broad categories (skin, neurological, respiratory, enteric). For each patient, at least one sequence of the major capsid protein VP1 gene has to be provided of the agreed genomic region (nucleotides 2,604–2,909 NC_001612, CVA16). In addition, samples that could not be typed as an enterovirus but were typed as poliovirus-like, were sent to the enterovirus section of the Center for Infectious Disease Control at RIVM, as part of the enterovirus surveillance programme in place, to document the absence of wild-type poliovirus circulation.

Data sharing and confidentiality agreement

Participants worked with a confidentiality agreement, consenting to the use of the data to provide surveillance overviews and alerts and to the right to publish the data, with proper acknowledgement, in case of public health emergencies. All participants can access and download the data, but they cannot be used without the consent of the data provider.

Enterovirus diagnostics and sequencing

Each laboratory used a laboratory-developed test, adapted from the protocol described by Nix et al. [10] (2006) for the detection of enteroviruses. One laboratory used an additional protocol described by McWilliam Leitch et al. [11] for cerebrospinal fluid samples. All laboratories participated in an external proficiency testing programme organised through Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom, an International Organization of Standardization (ISO) 17043-accredited organisation. Amplification of the 5' non-coding region of enterovirus was performed at the individual participating laboratory.

Genotype assignment using a standardised sequence-based typing tool

Upon entering of sequences into the database, an automated algorithm was run to assign the genotype. This tool has been validated against most currently known picornaviruses and has been shown to correlate highly with the serotype assignment [8].

Results

Questionnaire information

In addition to enterovirus, the seven participating laboratories indicated that they performed systematic genotyping for influenza virus (n=7), hepatitis B virus (n=6) and hepatitis C virus (n=5), primarily related to monitoring of treatment. Some laboratories also typed parechoviruses (n=5), rhinoviruses (n=3), hepatitis E virus (n=3), norovirus (n=2), hepatitis A virus (n=2), cytomegalovirus (n=2), herpes simplex virus (HSV) (n=2), adenovirus (n=2), human immunodeficiency virus (HIV) (n=3), as well as hepatitis B virus and hepatitis C virus for specific research or clinical study-related questions. A need for a more structured collaboration between the laboratories, possibly including the operation of a joint reference database, was indicated by the majority of respondents regarding influenza virus, parechovirus, rhinovirus and hepatitis B virus. For the less commonly used typing approaches, a need for collaboration was expressed for hepatitis viruses A, C and E. Given the consensus that a type of collaborative network would meet a need, a pilot TYPENED database was set up for enteroviruses.

Pilot enterovirus database

As of 1 May 2012, a total of 651 human enterovirus (HEV) sequences were submitted to the TYPENED database, representing all enterovirus-positive clinical samples that were successfully sequenced at six of the collaborating laboratories from 1 January 2010 to 31 December 2011. Most of the sequences belonged to HEV-A (n=168; 25.8%) and B (n=466; 71.6%), whereas only a few belonged to HEV-C (n=6; 0.9%) and D (n=6; 0.9%). Following automatic typing of the sequences submitted

to the TYPENED database, it appeared that some of the viruses that were enterovirus positive in the molecular diagnostic assay appeared to be a rhinovirus A (n=5; 0.8%), most probably due to the cross-reactivity of the primers used for detection. In addition, three poliovirus sequences were identified within the HEV-C set: all three isolates were obtained from children from the former Netherlands Antilles (Curaçao and Sint Maarten), where oral polio vaccines were used.

The laboratories that submitted the sequences received samples from laboratories all over the Netherlands. Although the numbers per serotype were not always very large, some clusters of serotypes over time could be observed (detailed data not shown). For example, of the 48 CV-A9 sequences submitted, 43 were found in samples collected from May to August 2010 with a clear peak (n=38) in June and July. In addition, five of the six EV-D68 sequences were found in samples collected from August to November 2010; 46 of the 65 E-7 sequences were found in samples collected from May to August 2011 and 51 of the 69 E-25 sequences were found in samples collected from August to December 2011.

Discussion

We have described a data-sharing concept that combines the capacities of clinical and public health laboratories in the Netherlands in a database to which all laboratories have equal and full access. After initial discussions to align expectations and develop a code of conduct, all laboratories were able to share a first set of historical data within two months. One of the triggers for the development of this concept was the concern that current enterovirus surveillance which is based on cell culture isolation is no longer the preferred method for enterovirus detection at hospital level and information obtained through other typing methods would not be captured centrally [12].

We managed to get consensus on the typing protocol and a data sharing agreement between the central public health laboratory (RIVM), large university laboratories and some large general hospitals that are geographically dispersed, thus potentially enabling broad coverage of surveillance of viruses of common interest. Within the enterovirus pilot, all sequences generated in two years by six of the seven collaborating laboratories were shared.

One pitfall of a consensus typing method may be that some viruses will be missed if they are not detected in the particular molecular test. This is of concern, given that the previously common practice of viral culture, which could serve as a safety net, is diminishing very rapidly. Most laboratories maintain these culture facilities only to grow control material for molecular assays. Since RNA viruses diverge rapidly, there is a need to get updated full-length sequences, not only for epidemiological reasons but also to keep diagnostic assays based on molecular testing up to date. At present, the availability of whole genome sequences is limited, but with next generation sequencing techniques rapidly coming within reach of academic and even clinical laboratories, this situation will change quickly.

The same system is currently being set up for a number of other viruses for which collaboration was valued according to the questionnaire – with parechovirus, norovirus and hepatitis E virus on the priority list [13-15]. Sequence-based characterisation is becoming more common within the larger diagnostic centres: the availability of sequence-based information will assist both the clinicians and diagnostic laboratories as well as the public health laboratories.

The concept of TYPENED in the Netherlands has been shown to be an effective means of close collaboration and the participating laboratories are willing to extend this collaboration to other targets. Furthermore, by using sequencing technologies, a more in-depth analysis of circulating strains can be carried out, as individual sequences can be analysed, instead of serotypes. Sequences have a much higher discriminatory power, as most sequences within one serotype will be different from each other, thus facilitating, for example, the tracing of transmission patterns. Sequence techniques are particularly valuable for viruses that are difficult to grow. In an economic climate with shrinking budgets, it may prove difficult for facilities to perform sequencing for diagnostic and epidemiological purposes, although it is expected that large centres will continue to perform routine sequencing. The TYPENED model seeks to maximise the use of data generated both in clinical and public health laboratories, for clinical care and for surveillance purposes. The harmonisation of typing protocols and sharing of data with a more extensive group of laboratories, or even cross-border centres, will be a next step.

References

1. Bean NH, Martin SM. Implementing a network for electronic surveillance reporting from public health reference laboratories: an international perspective. Emerg Infect Dis. 2001;7(5):773-9.
2. Emori TG, Gaynes RP. An overview of nosocomial infections, including the role of the microbiology laboratory. Clin Microbiol Rev. 1993;6(4):428-42.
3. Workman MR, Wall PG, Tearle P, O'Mahony M, Brunton WA. Active surveillance of health and safety in microbiology laboratories. Commun Dis Rep CDR Rev. 1995;5(4):R54-6.
4. Deyde VM, Sampath R, Gubareva LV. RT-PCR/electrospray ionization mass spectrometry approach in detection and characterization of influenza viruses. Expert Rev Mol Diagn. 2011;11(1):41-52.
5. Endimiani A, Hujer KM, Hujer AM, Kurz S, Jacobs MR, Perlin DS, et al. Are we ready for novel detection methods to treat respiratory pathogens in hospital-acquired pneumonia? Clin Infect Dis. 2011;52 Suppl 4:S373-83.
6. Fournier-Wirth C, Coste J. Nanotechnologies for pathogen detection: Future alternatives? Biologicals. 2010;38(1):9-13. Epub 2010 Jan 15.
7. Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, et al. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg Infect Dis. 2012;18(11):e1.
8. Kroneman A, Vennema H, Deforche K, v d Avoort H, Peñaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. J Clin Virol. 2011(2):121-5.
9. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41(D1):D36-42.
10. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. J Clin Microbiol. 2006;44(8):2698-704.
11. McWilliam Leitch EC, Harvala H, Robertson I, Ubillos I, Templeton K, Simmonds P. et al. (2009). Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. J Clin Virol. 44(2):119-24.
12. Manzara S, Muscillo M, La Rosa G, Marianelli C, Cattani P, Fadda G. Molecular identification and typing of enteroviruses isolated from clinical specimens. J Clin Microbiol. 2002;40(12):4554-60.
13. Wolthers KC, Benschop KS, Schinkel J, Molenkamp R, Bergevoet RM, Spijkerman IJ, et al. Human parechoviruses as an important viral cause of sepsislike illness and meningitis in young children. Clin Infect Dis. 2008;47(3):358-63.
14. Nelson KE, Kmush B, Labrique AB. The epidemiology of hepatitis E virus infections in developed countries and among immunocompromised patients. Expert Rev Anti Infect Ther. 2011;9(12):1133-48.
15. Verhoef L, Williams KP, Kroneman A, Sobral B, van Pelt W, Koopmans M; et al. Selection of a phylogenetically informative region of the norovirus genome for outbreak linkage. Virus Genes. 2012;44(1):8-18.

Imported diphyllobothriasis in Switzerland: molecular methods to define a clinical case of *Diphyllobothrium* infection as *Diphyllobothrium dendriticum*, August 2010

F de Marval[1], B Gottstein[2], M Weber[3], B Wicht (barbarawicht@yahoo.it)[4]
1. Dianalabs, Geneva, Switzerland
2. Institute of Parasitology, University of Bern, Bern, Switzerland
3. Private practice, Chêne-Bourg, Switzerland
4. Istituto cantonale di microbiologia, Bellinzona, Switzerland

Citation style for this article:
de Marval F, Gottstein B, Weber M, Wicht B. Imported diphyllobothriasis in Switzerland: molecular methods to define a clinical case of Diphyllobothrium infection as Diphyllobothrium dendriticum, August 2010 . Euro Surveill. 2013;18(3):pii=20355. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20355

Article submitted on 07 May 2012 / published on 17 January 2013

Following a first clinical case of infection by *Diphyllobothrium dendriticum* in Switzerland in 2006, we report a second case in the country. The species was identified by molecular methods. In the Swiss, French and Italian subalpine regions, human diphyllobothriasis has seen a comeback since the late 1980's, and *Diphyllobothrium latum* is usually considered the causative agent of the disease. In addition, several locally acquired and imported clinical infections due to allochthonous *Diphyllobothrium* species have been documented in the last years. Due to the colonisation potential of these parasites and their probably underestimated presence in the human population, there is a need for discriminating them at the medical laboratory level. Because the morphological characters are very similar among the different taxa, a correct identification requires the use of molecular methods. Molecular identification would improve diagnosis and help monitor the distribution of *Diphyllobothrium* species in Europe.

Introduction

Diphyllobothriasis is due to tapeworms of the genus *Diphyllobothrium* and is acquired through the consumption of raw or poorly cooked freshwater fish. In distinct stages of their life cycle, *Diphyllobothrium* spp. occupy different host species. The eggs present in the water are ingested by small crustaceans, such as copepods, constituting the first intermediate hosts in which they develop to a larval stage. As crustaceans are predated by fish, fish become the second intermediate hosts where *Diphyllobothrium* larvae further develop. The definite hosts are fish eating mammals (including humans) or birds, where *Diphyllobothrium* matures into an adult stage.

Human diphyllobothriasis is often a mild illness: about half of the patients are asymptomatic, while the others mostly suffer from minor discomforts, such as diarrhoea, abdominal pain and digestive troubles. Rare cases of megaloblastic anaemia associated with vitamin B12 deficiency have been reported in the literature [1].

Globally, the incidence of human diphyllobothriasis has decreased in the last 20 years, particularly in northern and eastern Europe [2]. Nevertheless, in some Swiss, French and Italian subalpine regions the disease has seen a comeback, as shown by the more than 530 cases reported since 1987 from around lakes Leman (Geneva), Morat, Bienne, Maggiore, Lario (Como), Iseo and Garda [2]. In these areas, the species *Diphyllobothrium latum* is considered the causative agent of diphyllobothriasis. However, infections due to allochthonous species (*D. nihonkaiense, D. dendriticum*) have been recently documented [3-7]. A first clinical case due to *D. dendriticum* was diagnosed in 2006 in Switzerland, raising the question of potential transmission to susceptible intermediate hosts present in the local environment [4]. In this report, we describe a second case of symptomatic infection by *D. dendriticum* in Switzerland. Confirmation of the species was done by molecular identification.

A four year-old boy expelled tapeworm segments in stool, 12 days after returning from a 15-day holiday travel in Singapore and Bali in August 2010. He had been suffering from abdominal pains (cramps) and loose stools since his return to Switzerland. Standard laboratory procedures held in the clinical laboratory (Dianalabs) led to the identification of *Diphyllobothrium* spp., based upon the presence of typical operculated eggs in segments. However, because of the unusual shape of a tapeworm proglottid (longer than wide as if stretched, with an excentred uterus) and the possible Asian origin of infection, the specimen was preserved in 70° ethanol and sent to the Institute of Parasitology in Bern and to the Cantonal Institute of Microbiology in Bellinzona, where it was identified as *D. dendriticum* by molecular methods.

Mebendazole was first administered to the patient, with no curative effect, as confirmed by the persisting presence of eggs in stool after three weeks. The patient was then re-medicated with praziquantel and recovered promptly. No parasites were found upon stool testing six weeks after praziquantel therapy.

The patient's family did not present with symptoms and underwent no further investigation, except for the seven year-old patient's sister who was checked for intestinal parasites but found negative.

Methods

The faecal specimen was processed by standard sedimentation technique [8] to concentrate putatively present *Diphyllobothrium* eggs and subsequently assess these by light microscopy. A segment of proglottids of approximately 5 cm length was processed for staining with lacto-acetic carmine according to Rukhadze and Blajin [9].

Genomic DNA from about 50 mg of proglottid tissue was extracted with the DNeasy Blood and Tissue Kit (Qiagen). Polymerase chain reaction (PCR) was performed using the Taq PCR Master Mix Kit (Qiagen) with primers targeting a region of the 5.8S ribosomal RNA (5.8S rRNA) comprising internal transcribed spacers (ITS) 1 and 2 [10], the 18S ribosomal RNA (18S rRNA) [11] and the cytochrome c oxidase subunit 1 gene (cox1) [3,12] sequences. The amplification of all targets was carried out under the following conditions: 5 min at 94 °C, 35 cycles consisting of 30 s at 94 °C, 40 s at 45 °C, 1 min at 72 °C, and a final extension step of 10 min at 72 °C. Amplicons were visualised by electrophoresis in a 0.8% agarose gel containing ethidium bromide, and purified through Sephadex G-50 columns (GE Healthcare). DNA was quantified with a ND-100 Spectrophotometer (NanoDrop Technologies Inc.). Sequencing was performed with the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems), according to the provider's recommendations. Samples were purified by osmosis with 0.025 µm nitrocellulose filters (Millipore) in tris ethylenediaminetetraacetic acid (TE) buffer pH 8 for two hours. Eight µl of purified solution were placed in 0.5 ml Genetic Analyzer Sample Tubes with 12 µl Hi-Di Formamide (Applied Biosystems). Samples were then loaded in an automated sequencing system (ABI PRISM 310 Genetic Analyzer; Perkin Elmer).

Sequence electropherograms were corrected by using the software EditSeq (DNASTAR Inc.). Their identity was first checked by basic local alignment search tool (BLAST) [13]. Sequence fragments of 657 and 375 nucleotides in length, derived from the PCRs targeting the ITS1-5.8SrRNA and cox1 genes were then respectively compared to representative ITS1-5.8SrRNA or cox1 sequences from different *Diphyllobothrium* spp. by pairwise and multiple alignments using ClustalW [14] with the software Molecular Evolutionary Genetics Analysis (MEGA) version 4.0 [15]. Phylogenetic trees (neighbour-joining method; Kimura-2 parameters; bootstrap test for 500 replicates) were subsequently inferred from the alignments.

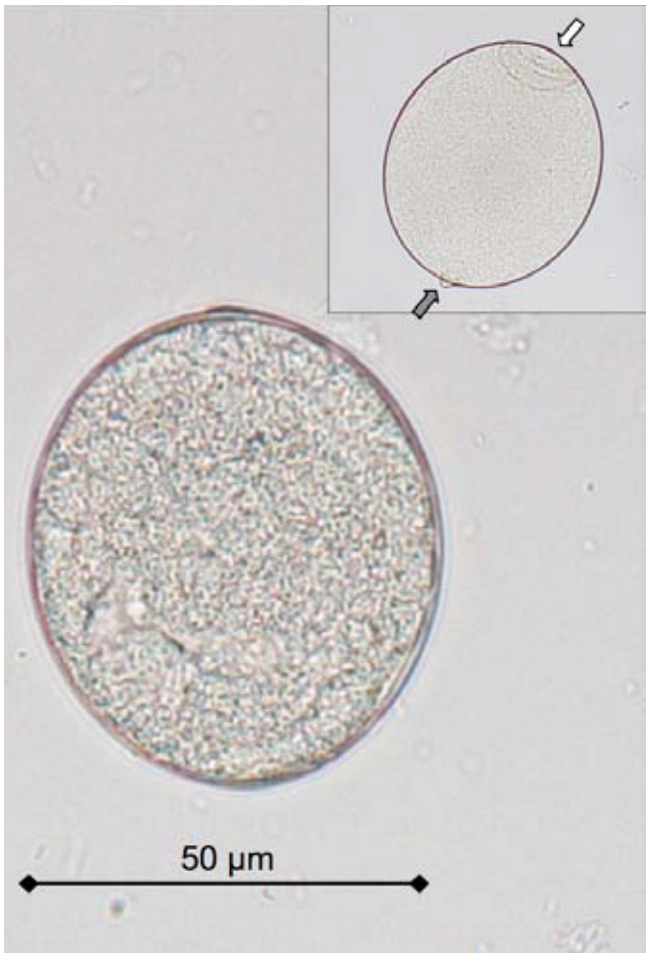
Results

Microscopical analyses of the coprological sediment revealed the presence of oval-shape unembryonated eggs (mean size: 49 x 64 µm; range: 48.5–52.5 x 62.5–70 µm), characterised by the presence of a hardly visible operculum and a small knob at the abopercular end (Figure 1).

Microscopical analyses of the stained proglottids revealed the presence of only one set of reproductive organs per proglottid (Figure 2). The central uterine structure showed several rosette-shaped loops. Morphological criteria matched to those described

Figure 1

Diphyllobothrium dendriticum eggs recovered from a patient stool, Switzerland, 2010

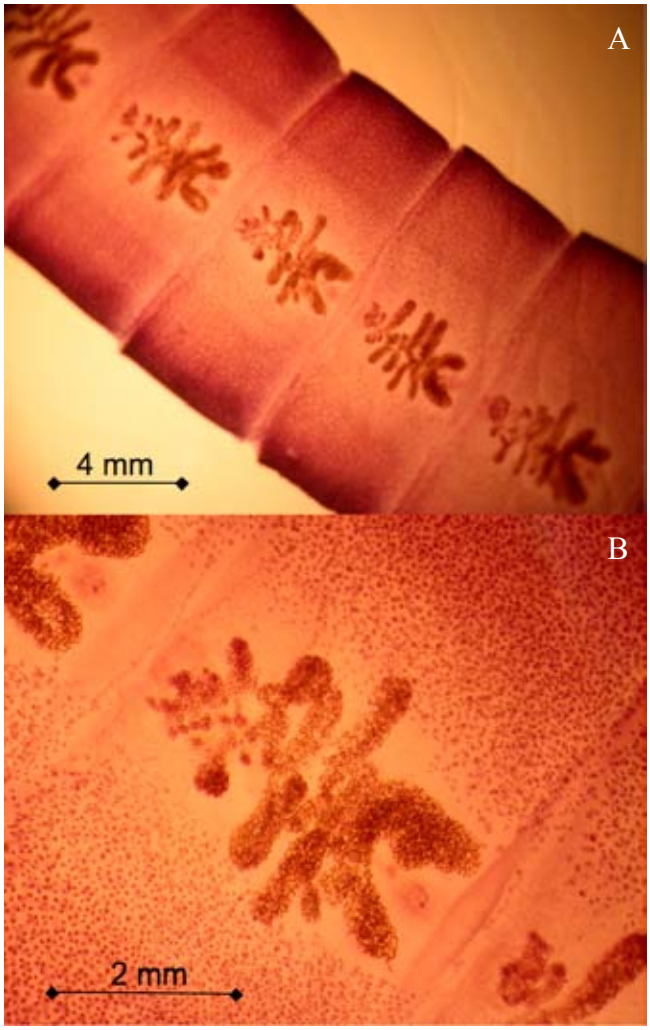


White arrow: operculum; grey arrow: knob.

The pictures were taken under 400x magnification and the mean size of the eggs was 49 x 64 µm.

Figure 2

Diphyllobothrium dendriticum segment recovered from a patient stool, Switzerland, 2010



A. *Diphyllobothrium dendriticum* segment.
B. Enlargement of an individual proglottid where the central uterine structure forms several rosette-shaped loops.

for *Diphyllobothrium* spp. [16], while a species-specific identification was not possible based upon morphology.

The nucleotide sequences of the ITS1-5.8S rRNA, 18S rRNA and cox1 derived from the *Diphyllobothrium* affecting the patient were deposited in public databases and are available in the European Molecular Biology Laboratory (EMBL), GenBank and DNA Data Bank of Japan (DDJB) databases under accession numbers HQ682065, HQ682066 and HQ682067. Due to technical problems, the sequence of ITS2 was not fully obtained and therefore not used in the analysis.

The results of BLAST search showed that all the sequenced targets respectively reached ≥99% identity (highest scores) at the nucleotide level with *D. dendriticum* homologous reference sequences of respective GenBank accession numbers FM204787 (ITS1-5.8S rRNA), DQ768164, DQ181945 (18S rRNA) and AM412738 (cox1). The 18S rRNA sequence also showed 99% identity with those of *D. ditremum* (GenBank accession numbers: DQ768165, DQ181944) and *D. latum* (GenBank accession number: DQ316795).

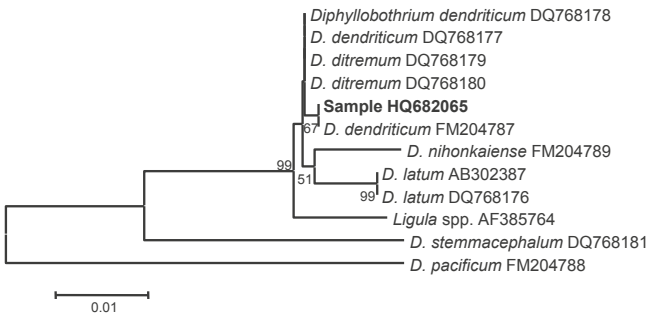
The position of the sample sequences of ITS1-5.8S rRNA and cox1 regions in the phylogenetic trees confirmed the *Diphyllobothrium* spp.affecting the patient as *D. dendriticum* (Nitsch, 1824) (Figures 3 and 4). The phylogenetic tree built on the basis of 18S rRNA sequences is not shown, because this target is not useful for the discrimination of *D. dendriticum* from other species [2].

Discussion

At the time of tapeworm evacuation, the patient was known to have had a meal in a Japanese restaurant in Singapore (on day 4 of his journey) where the family had consumed various dishes (fish, chicken). According to the recollection of the family members, the meal did not contain raw fish. Retrospective investigations revealed that the patient regularly consumed fish, e.g. smoked salmon with pasta at home. This salmon was always bought in the same department store in France and was of the same brand. According to the product information, it belonged to the species *Salmo salar,* was farmed in Norway, smoked in France and guaranteed 'never frozen'. The homemade sauce was made by dropping slices of smoked salmon into boiling cream

Figure 3

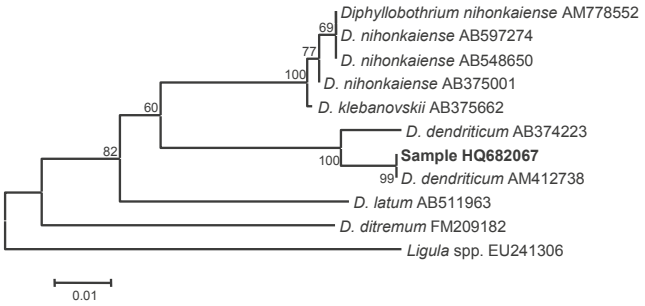
Neighbour-joining tree based on 5.8S ribosomal RNA internal transcribed spacer 1 sequences, to determine the *Diphyllobothrium* species affecting a patient, Switzerland, 2010



The tree is based on 5.8S ribosomal RNA internal transcribed spacer 1 (5.8S rRNA-ITS1) sequence fragments of 657 bp. The 'sample' refers to the sequence of the unknown *Diphyllobothrium* spp. affecting the patient. GenBank accession numbers of all the sequences used to construct the tree are indicated. On the tree nodes, Kimura-2 parameters bootstrap values >50 for 500 replicates are shown.

Figure 4

Neighbour-joining tree based on cytochrome c oxidase subunit 1 sequences, to determine the *Diphyllobothrium* species affecting a patient, Switzerland, 2010



The tree is based on cytochrome c oxidase subunit 1 (cox1) sequence fragments of 375 bp. The 'sample' refers to the sequence of the unknown *Diphyllobothrium* spp. affecting the patient. GenBank accession numbers of all the sequences used to construct the tree are indicated. On the tree nodes, Kimura-2 parameters bootstrap values >50 for 500 replicates are shown.

and leaving to cook for one minute before serving. According to his mother, the patient enjoyed picking pieces of uncooked fish during the preparation of the meal. Short before travelling to south-east Asia, the patient consumed poorly cooked perch fillets (*Perca fluviatilis*) fished in the Leman lake in a restaurant near home. He also ate sushi in Switzerland one week before tapeworm segments were noticed.

Aquatic birds (especially Laridae) are usually the definitive hosts of *D. dendriticum*, while humans are only occasionally infected. According to literature, the incubation period of *D. dendriticum* in herring gulls (*Larus argentatus*) lasts from five to 20 days [17]. In humans, *Diphyllobothrium* plerocercoids generally develop into mature adults in two to six weeks [2].

Based on these observations, the source of the patient's infection might be either salmon from Norway, unknown fish from Asia or perch from Switzerland. However, because the presence of *D. dendriticum* in perch has not been documented so far, it is unlikely that the perch meal was the source of infection. The sushi meal in Switzerland can be excluded because the symptoms were already present at that time.

In the last two decades, diphyllobothriasis has shown a recrudescence in a number of European countries, especially in the subalpine lakes region [2]. The use of molecular methods also showed the presence of allochthonous *D. nihonkaiense* and *D. dendriticum* in France, Switzerland, Finland and the Czech Republic (Table). Except for two cases of *D. dendriticum* diagnosed in Switzerland [4] and the Czech Republic [18] that were probably acquired abroad (Norway, Alaska or Canada), most of the documented infections were locally acquired, attributed to imported salmons.

Susceptible intermediate hosts for *D. dendriticum* such as copepods (*Eudiaptomus* and *Cyclops species* [19]) and fish (*Oncorhynchus mykiss, Salmo trutta, Coregonus clupeaformis, C. albula, C. lavaterus* [19-21]) are known to be present in Switzerland. This would theoretically allow the introduction and the autochtonous transmission of the parasite. This colonisation potential emphasises the need for correct identification of *Diphyllobothrium* species involved in clinical cases of infection.

Interestingly, in the two Swiss cases due to *D. dendriticum*, the molecular investigation was undertaken because of the unusual shape of some segments.

The transmission of *Diphyllobothrium* plerocercoids may be prevented by freezing fish at -20 °C for one to seven days, depending on its thickness. The Swiss law [22] provides that it is forbidden to sell any kind of fresh fish products (both local and imported) intended to be consumed raw or semi-raw, unless they have been stored at a temperature not higher than -20 °C for a least 24 hours. However, it has been demonstrated that this rule is not always followed properly [23], and of course there is no control of food bought abroad by individuals for personal use. Therefore cooking at 55 °C or more, for at least 5 min, remains the most reliable way to prevent transmission of parasitic worms possibly present in fresh fish [2].

Conclusion

A correct diagnosis has become crucial to evaluate the distribution of human-infecting *Diphyllobothrium* species, as well as their fish hosts, and to prevent the spread of allochthonous parasites in aquatic environments. Due to the difficulties in discriminating the different *Diphyllobothrium* taxa by morphological characters, molecular analysis has proven to be fundamental to identify these helminths at the species level. A cheap and rapid molecular test based on multiplex PCR with partial cytochrome *c* oxidase subunit 1 (cox1) gene, without the need of sequencing, was recently developed for the differential identification of the most common species infecting humans [24]. In case of atypical specimens of proglottids and eggs and/or specimens derived from patients who have been abroad, this test could be used to verify the parasite's identity. Improving the diagnosis of *Diphyllobothrium* parasites would help to monitor the distribution of species in Europe and trace the source of infections, an important goal at a time when eating habits are changing, fish markets are globalising and climate is changing [25].

Table

Case reports of allochthonous *Diphyllobothrium* infections in Europe, identified by molecular methods, 2005–2011

Year of parasite recovery	Country	Patient (age in years)	*Diphyllobotrhium* species	Parasite characteristics	Clinical features	Probable source of infection	Molecular identification	Therapy	Reference
2005	Finland	Man (60)	*D. nihonkaiense*	ND	ND	ND	Cox1	ND	Data not shown
	France	Woman (44)	*D. nihonkaiense*	Egg size 41 × 60 µm	Nausea, epigastric pain, diarrhoea	Wild salmon carpaccio (*Oncorhynchus keta*) imported from Canada (Gulf of Alaska, Pacific Ocean), purchased in France	Cox1, MT-ND3	Single dose of praziquantel at 10 mg/kg.	Yera et al., 2006 [6]
	Switzerland	Woman (55)	*D. nihonkaiense*	Egg size 56.68–57.82 × 43.58–45.42 µm	Diarrhoea (probably due to another cause); proglottids passed in faeces	Raw salmon and sushi purchased in Switzerland	18S rRNA, cox1, ITS1 and 2	Praziquantel	Wicht et al., 2007 [3]
2006	Switzerland	Woman (52)	*D. nihonkaiense*	ND	None	Raw Pacific salmon (*O. keta*) imported from Canada or North America (Pacific Ocean), purchased in France	18S rRNA, cox1, ITS1 and 2	Single dose of praziquantel	Wicht et al., 2007 [3]
	Switzerland	Boy (5)	*D. nihonkaiense*	Egg size 57.5–65.0 × 40.0–42.5 µm	Mild eosinophily (7.1%)	Pacific salmon purchased in Switzerland	ITS1, cox1, MT-ND3	Single dose of praziquantel at 12 mg/kg	Shimizu et al., 2008 [5]
	Switzerland	Woman (59)	*D. dendriticum*	Egg size 49.62–63.86 × 35.75–43.41 µm; some proglottids longer than wide, with excentred utera	Chronically relapsing courses of diarrhoea	Regular consumption of wild salmon, Japanese sushi or fish carpaccio; journeys in Canada, Alaska and Norway	18S rRNA, cox1	Single dose of praziquantel at 10 mg/kg	Wicht et al., 2008 [4]
2008	France	Woman (33)	*D. nihonkaiense*	Egg size 57 × 44 µm	Persistant, mild eosinophily	Wild salmon carpaccio or marinade purchased in France	Cox1	Single dose of praziquantel at 15 mg/kg	Paugam et al., 2009 [7]
2010	Switzerland	Boy (4)	*D. dendriticum*	Egg size 48.5–52.5 × 62.5–70 µm; one of the proglottids longer than wide, with excentred uterus	Abdominal cramps, loose stools	Smoked salmon (*Salmo salar*) imported from Norway, purchased in France; poorly cooked perch from Switzerland; fish eaten in Asia	ITS1 and 2, 18S rRNA, cox1	Praziquantel	Present paper
2011	Czech Republic	Woman (28)	*D. dendriticum*	Egg size 49.5 × 64 µm	No symptoms reported; proglottids passed in faeces	Salmons (*O. tshawytscha, O. keta, O. kisutch, O. nerka, O. gorbuscha*) and *Coregonus autumnalis* eaten in Alaska	Cox1	Single dose of praziquantel (750 mg)	Kuchta et al., 2012 [18]

Cox1: cytochrome c oxidase subunit 1; ITS: internal transcribed spacer of the 5.8S ribosomal ribonucleic acid; MT-ND3: mitochondrially encoded NADH dehydrogenase 3; NADH: reduced nicotinamide adenine dinucleotide; ND: not determined; rRNA: ribosomal ribonucleic acid.

Acknowledgments

We gratefully thank the patient's mother, who gave spontaneous and detailed information on the family eating habits, the probable sources of infection and her cooking recipes. Thanks are also due to Seppo Meri (Department of Bacteriology and Immunology, University of Helsinki, Finland) for providing the specimen of *D. nihonkaiense* found in Finland.

References

1. von Bonsdorff B. Diphyllobothriasis in man. New York: Academic Press; 1977.
2. Scholz T, Garcia HH, Kuchta R, Wicht B. Update on the human broad tapeworm (genus Diphyllobothrium), including clinical relevance. Clin Microbiol Rev. 2009;22(1):146-60.
3. Wicht B, de Marval F, Peduzzi R. Diphyllobothrium nihonkaiense (Yamane et al., 1986) in Switzerland: first molecular evidence and case reports. Parasitol Int. 2007;56(3):195-9.
4. Wicht B, de Marval F, Gottstein B, Peduzzi R. Imported diphyllobothriasis in Switzerland: molecular evidence of Diphyllobothrium dendriticum (Nitsch, 1824). Parasitol Res. 2008;102(2):201-4.
5. Shimizu H, Kawakatsu H, Shimizu T, Yamada M, Tegoshi T, Uchikawa R, et al. Diphyllobothriasis nihonkaiense: possibly acquired in Switzerland from imported Pacific salmon. Intern Med. 2008;47(14):1359-62.
6. Yera H, Estran C, Delaunay P, Gari-Toussaint M, Dupouy-Camet J, Marty P. Putative Diphyllobothrium nihonkaiense acquired from a Pacific salmon (Oncorhynchus keta) eaten in France; genomic identification and case report. Parasitol Int. 2006;55(1):45-9.
7. Paugam A, Yéra H, Poirier P, Lebuisson A, Dupouy-Camet J. Bothriocéphalose à Diphyllobothrium nihonkaiense: un nouveau risque lié à la consommation de saumon [Diphyllobothrium nihonkaiense infection: a new risk in relation with the consumption of salmon]. Presse Méd. 2009;38(4):675-7. French.
8. Rommel M, Eckert J, Kutzer E, Körting W, Schnieder T, editors. Veterinärmedizinische Parasitologie. 5th edition. Berlin: Paul Parey Verlag; 2000. German.
9. Rukhadze, Blajin. On a method for staining flukes and tapeworm segments as whole microscopical preparations. J Trop Med Hyg. 1929;33:342-5.
10. Luton K, Walker D, Blair D. Comparisons of ribosomal internal transcribed spacers from two congeneric species of flukes (Platyhelminthes: Trematoda: Digenea). Mol Biochem Parasitol. 1992;56(2):323-7.
11. Mariaux J. A molecular phylogeny of the Eucestoda. J Parasitol. 1998;84(1):114-24.
12. Bowles J, Blair D, McManus DP. Genetic variants within the genus Echinococcus identified by mitochondrial DNA sequencing. Mol Biochem Parasitol. 1992;54(2):165-73.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
14. Thompson JD, Higgins DG, Gibson TJ. Clustal W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673-80.
15. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24(8):1596-9.
16. Schmidt GD. How to know the tapeworms. Dubuque: WMC Brown Comp, Publishers; 1970.
17. Pronin NM, Timoshenko TM, Sanzhieva SD. Dynamics of egg production of the cestode Diphyllobothiurm dendriticum (Nitzsch, 1824) (Cestoda: Pseudophyllidea) and the concept of fecundity in helminths. Folia Parasitol (Praha). 1989;36(1):49-57.
18. Kuchta R, Wicht B, Ditrich O, Kubáčková P. Rare species of human broad tapeworms (Diphyllobothrium) in Europe. Oral presentation presented at: Xth Czech and Slovak Parasitological Days; 2012 May 28-June 1; Brno, Czech Republic.
19. Rahkonen R, Aalto J, Koski P, Särkkä J, Juntunen K. Cestode larvae Diphyllobothrium dendriticum as a cause of heart disease leading to mortality in hatchery-reared sea trout and brown trout. Dis Aquat Org. 1996;25:15-22.
20. Dick TA, Poole BC. Identification of Diphyllobothrium dendriticum and Diphyllobothrium latum from some freshwater fishes of central Canada. Can J Zool. 1985;63(1):196-201.
21. Dolezel, MC, Crompton DWT. Platyhelminth infections in powan, Coregonus lavaretus (L.), from Loch Eck and Loch Lomond, Scotland. Helminthologia. 2000;37(3):147-52.
22. Ordonnance du Département Fédéral de l'Intérieur sur l'hygiène (OHyg) du 23 novembre 2005 (état le 1er novembre 2010) n° 817.024.1, Art. 42 [Ordinance of the Federal Department of Home Affairs on Hygiene of 23 November 2005 (as of 1 November 2010) No. 817.024.1, Art. 42]. French. Available from: http://www.admin.ch/ch/f/rs/8/817.024.1.fr.pdf
23. Wicht B, Zali O. Contrôle parasitologique des poissons vendus dans les restaurants genevois. [Checking fish sold in Geneva restaurants for parasites]. Geneva; Service de la consommation et des affaires vétérinaires. Sep 2007-Oct 2008. p. 13. French.
24. Wicht B, Yanagida T, Scholz T, Ito A, Jiménez JA, Brabec J. Multiplex PCR for differential identification of broad tapeworms (Cestoda: Diphyllobothrium) infecting humans. J Clin Microbiol. 2010;48(9): 3111-6.
25. Torres P, Cuevas C, Tang M, Barra M, Franjola R, Navarrete N, et al. Introduced and native fishes as infection foci of Diphyllobothrium spp. in humans and dogs from two localities at lake Panguipulli in southern Chile. Comparative parasitology. 2004;71(2):111-7.

From theory to practice: molecular strain typing for the clinical and public health setting

R V Goering (rgoeri@creighton.edu)[1], R Köck[2], H Grundmann[3], G Werner[4], A W Friedrich[3], on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM)[5]

1. Department of Medical Microbiology and Immunology, Creighton University School of Medicine, Omaha, USA
2. Institute of Hygiene, University Hospital Münster, Münster, Germany
3. Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
4. National Reference Centre for Staphyloccoci and Enterococci, Robert Koch Institute, Wernigerode Branch, Wernigerode, Germany
5. European Society for Clinical Microbiology and Infectious Diseases, Basel, Switzerland

Citation style for this article:
Goering RV, Köck R, Grundmann H, Werner G, Friedrich AW, on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM). From Theory to Practice: Molecular Strain Typing for the Clinical and Public Health Setting. Euro Surveill. 2013;18(4):pii=20383. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20383

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

The persistence and transmission of infectious disease is one of the most enduring and daunting concerns in healthcare. Over the years, epidemiological analysis especially of bacterial etiological agents has undergone a remarkable evolutionary metamorphosis. While initially relying on purely phenotypic characterisation, advances in molecular biology have found translational application in a number of approaches to strain typing which commonly centre either on 'epityping' (molecular epidemiology) to characterise outbreaks, perform surveillance, and trace evolutionary pathways, or 'pathotyping' to compare strains based on the presence or absence of specific virulence or resistance genes. A perspective overview of strain typing is presented here considering the issues surrounding analyses which are employed in the localised clinical setting as well as at a more regional/national public health level. The discussion especially considers the shortcomings inherent in epidemiological analysis: less than full isolate characterisation by the typing method and limitations imposed by the available data, context, and time constraints of the epidemiological investigation (i.e. the available epidemiological window). However, the promises outweigh the pitfalls as one considers the potential for advances in genomic characterisation and information technology to provide an unprecedented aggregate of epidemiological information and analysis.

Introduction

Since the time of Semmelweis and Koch's Postulates, medical science has recognised the cause-and-effect relationship between the transmission of etiological agents and the persistence and spread of infectious disease. In this context, routine clinical and infection control interests commonly centre on the detection of multifocal patient infection or dissemination within a defined patient population (e.g. outbreak identification, control, or other rather short-term epidemiological issues). Conversely, public health concerns include local, regional, national, and international emergence and spread of pathogens, global microbiological and molecular surveillance, as well as longer term evolutionary interrelationships. Classical epidemiology uses the three parameters (time, place, person) to find epidemiological links. However, in both healthcare and community-associated infections today, those three parameters do not necessarily provide the desired resolution to identify an outbreak event or the causing pathogen. Clinical microbiology provides species-level isolate identification and molecular analysis provides the strain type or subtype fingerprint. Bringing these five parameters together provides the greatest hope of associating outbreaks of infectious disease with certain types of the same bacterial species. This perspective overview considers the epidemiological analysis of infectious diseases in both the clinical and public health setting, focusing on bacterial etiologies to illustrate issues associated with moving molecular strain typing from theory to practical application. Regardless of the setting, the interrelationships that strain typing seeks to clarify are generally in the context of epityping (i.e. transmission investigation (e.g. outbreak)) or pathotyping to compare strains based on the presence or absence of specific virulence genes. The former is emphasised here and discussed in the context of two principal challenges independent of the methods employed: isolate characterisation and the available data, context, and time constraints of the epidemiological investigation (i.e. the available epidemiological window).

The challenge of isolate characterisation

In both the clinical and public health setting, the assessment of potential interrelationships between isolates is based on a comparison of specific characteristics which ideally will identify (i.e. fingerprint) transmitted strains as the same type while not overlooking

epidemiologically relevant variants (subtypes) or mistakenly including unrelated isolates (i.e. issues of sensitivity and specificity). Isolate characterisation has been historically based on phenotypic assessment which is most certainly still of value (e.g. antibiograms, serotyping). However, recognition of the bacterial chromosome as the fundamental molecule of cellular identity has firmly established the importance of molecular (genomic) epidemiological evaluation. Thus, molecular approaches to isolate characterisation are considered here. In general, historical review reveals a consistent 'translational' trend of genotypic methods moving from the basic science laboratory to clinical application. These approaches to molecular epidemiology are reviewed more completely elsewhere [1,2] and are only summarised here to note the challenges faced in terms of providing definitive isolate characterisation for epidemiological purposes.

Simply stated, when it comes to epidemiological sensitivity and specificity the key methodological issues are: (i) the degree to which the targets/markers being analysed provide epidemiologically relevant information and (ii) the precision with which the queried characteristic(s) are identified and analysed. The former relates to epidemiological validation which has been considered elsewhere [3] and is beyond the scope of this discussion. However, by way of summary it is important to note that, regardless of analytical precision, other than whole genome sequencing (WGS) all methods strive to assess isolate interrelatedness based on a subset of targets that represent a genomically incomplete, but epidemiologically relevant, dataset. Thus, for these approaches, additional data is more informative than less (e.g. see [4]). In terms of precise data output, while newer methods employ instrumentation (e.g. capillary electrophoresis using an automated DNA sequencer [5]), a significant number of currently used protocols rely on visual inspection of data output generated by agarose gel electrophoresis (Table). While such analysis can be accurate for protocols involving the presence or absence of end point polymerase chain reaction (PCR) products, visual assessment of fragment-size comparisons (e.g. by agarose gel electrophoresis) can be problematic. For example, digestion of total cellular DNA by common restriction enzymes (restriction endonuclease analysis (REA)) can generate greater than 600 fragments from a typical 2 to 3 Mb bacterial chromosome. In addition, there is an element of imprecision in the visual comparison of DNA banding patterns in electrophoresis gels since DNA fragments differing by ±10% may be seen as identical [6]. This could amount to a 70 kb discrepancy, for example, in a pulsed-field gel with bands ca. 700 kb in size.

As noted earlier, the chromosome is the most fundamental molecule of identity in the cell. Thus, it is the sequence-based methods that ultimately hold the greatest promise for accurately assessing epidemiological interrelationships in problem pathogens. Reviewed elsewhere [2,7] these methods can be found in three general iterations: single locus sequence typing (SLST), multilocus sequence typing (MLST), and WGS (Table). Of these, the first two have found broad epidemiological application although, as noted above for other methods, both represent a genomically incomplete dataset, while WGS holds clear promise for providing total chromosomal analysis. While WGS was impossible with older dideoxy/chain termination sequencing technology [8], newer (i.e. next generation sequencing, NGS) methods have made this goal a reality. The technology behind NGS is discussed in detail elsewhere [7,9], however, from a strain typing

Table

Characteristics of methods commonly used for molecular epidemiology

Data generation	Chromosomal target(s)	Data output	Method examples
Restriction enzymes	Common restriction sites	DNA fragments visualised after agarose gel electrophoresis (AGE)	Restriction endonuclease analysis (REA)
Restriction enzymes	Common restriction sites	Ordered sequence scaffolds identified via instrument software	Optical mapping
Restriction enzymes	Rare restriction sites	DNA fragments visualised after AGE	Pulsed-field gel electrophoresis (PFGE)
Polymerase chain reaction (PCR)	Repetitive element or variable-number tandem repeat (VNTR) sequences	Amplified DNA fragments either visualised after AGE or via instrument software	Repetitive-element PCR (rep-PCR); VNTR typing; PCR ribotyping
DNA probes	Multiple genes	Hybridisation signal either identified visually or via instrument software	Microarray
DNA sequencing	Single or multiple genes	DNA sequence obtained via instrument software	*Staphylococcus aureus* protein A gene (spa) typing; multilocus sequence typing (MLST)
DNA sequencing	Whole genome	DNA sequence obtained via instrument software	Whole genome sequencing (WGS); next generation sequencing (NGS)

A full description of methods is reported elsewhere [1,2].

standpoint it is important to note that revolutionary developments in NGS have made WGS possible with benchtop instrumentation such as the Ion Torrent PGM (Life Technologies, Guilford), GS Junior (454 Life Sciences/Roche, Branford), and the MiSeq (Illumina, San Diego). Such instrumentation now allows WGS to be completed in hours to days with extensive multi-fold coverage allowing isolates to be compared down to the level of single nucleotide polymorphisms (SNPs). However, as with previous sequencing iterations, the critical issues for NGS are throughput, quality, read length and cost. All of these are currently in a state of flux as commercial technology improves and positions itself in the scientific marketplace. In addition, it must be noted that the present state of WGS has not reached accurate base-by-base total origin-to-termini output. For example, the assembly and analysis of the relatively short read lengths from current NGS platforms are problematic for repeat sequences (e.g. clustered regularly interspaced short palindromic repeats (CRISPRs), homopolymers, and variable-number tandem repeats (VNTRs) [10]). An additional bottleneck is the bioinformatics requirement for proper WGS annotation and analysis which at present is far from routine, with costs (in time and money) that may exceed that of the sequencing itself [11,12]. Nevertheless, these are exciting 'problems' to have, confirming that the scientific stage is clearly set for remarkable developments in this most fundamental approach to determining isolate epidemiological interrelationships.

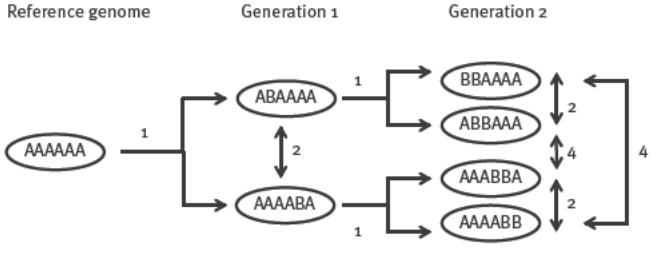
The challenge of the epidemiological window and detecting significant difference

Regardless of the epidemiological approach, the focus ultimately becomes data interpretation. Thus, it is important to note that while the term 'molecular' epidemiology implies a precise process, this is not always the case regardless of the method employed since epidemiological analysis always has an unavoidable context and time-driven component. A variety of environmental factors as well as interaction between the host and infectious agent may all influence the course of disease transmission. In addition, the time leading up to, as well as that required for, the epidemiological investigation provides opportunity for the outbreak strain to evolve. Whether in a clinical or public health setting, infectious disease scenarios benefiting from epidemiological evaluation do not typically give advance warning. Hence, in many investigations where the starting point of the epidemiological scenario (e.g. the source case or the outbreak source) is not identified, the process of data analysis attempts to work backward in time which, depending on the available information, may necessitate drawing conclusions based on probabilities rather than absolute certainty [13]. However, as with classical epidemiological approaches, molecular epidemiological analysis may to some extent implicate the source 'beyond a reasonable doubt'.

In the absence of a source isolate, all strain typing methods are challenged as the opportunity for chromosomal change over time increases the potential for genetic distance between epidemiologically related isolates (i.e. confounding the recognition of interrelationships in the isolates being analysed). This can be illustrated (Figure) considering a simple example of six epidemiologically-relevant characters ('A') in a reference genome (e.g. the characters could be restriction sites, specific genes, other chromosomal loci). Evolution through two generations, with sequential genetic events of unknown complexity (e.g. insertions, deletions, rearrangements, recombination) designated as changes from 'A' to 'B', results in second-generation genomes varying from each other by four differences. As the process continues through subsequent generations additional complexity in the population dramatically increases. This scenario illustrates the issue central to the interpretation of any bacterial strain typing data, the definition and detection of significant difference. This relates to the issues of sensitivity and specificity previously addressed, in particular specificity, which is important to insure adequate case definitions for outbreak investigations, in order to avoid inclusion of non-cases and detect maximum epidemiological associations between the isolates. Thus, for optimum epidemiological outcome, proper analysis of strain typing data requires knowledge of: (i) the genetics of the microbial pathogen (e.g. clock speed/rate of change of the characteristics being analysed), (ii) the limitations of the typing method, (iii) the degree of concordance between different typing methods, if more than one technique is applied in parallel, and (iv) the setting within which the issue is being studied. Regardless of the typing approach, these details must be considered in attempting to discern the relatedness and transmission patterns of infectious agents in both the clinical and public health setting.

Figure

Diagrammatic illustration of interrelationships between a reference genome and two subsequent generations each of which differs from the previous by a single genetic event



The reference genome has six epidemiologically-relevant characteristics (designated 'A'). Each generation differs from the previous by a single genetic event (indicated by the number 1 above the horizontal arrows) changing characteristics from A to B. For each generation, the numbers of genetic differences between members are indicated by figures on the side of the vertical arrows (adapted from [13]).

The 'typing Esperanto'

It is of utmost importance, that typing methods produce data that can be compared not only within the same laboratory or clinical setting, but also between different facilities. Therefore, the 'typing Esperanto' or language should produce data that are clear, reproducible, and include strain nomenclature which allows for the independent identification of specific types. However, it is important to note that the probability of an outbreak due to a certain strain type depends on its frequency in the associated environment (e.g. both within and outside of the healthcare setting, the community). The less frequent a strain type is, the more probable it becomes that multiple isolates (a cluster) of a certain strain type represent a true outbreak. Thus, epidemiological analysis must recognise the nuances associated with disease transmission such as distinguishing outbreaks from pseudo-outbreaks [14]. The latter occur frequently in environments associated with an endemic prevalence of antibiotic-resistant microorganisms. For example, in a clinical setting, patients on the same hospital ward may carry similar but distinct problem pathogens which could superficially mimic an outbreak. Useful typing should properly identify such a pseudo-outbreak thus helping to avoid inappropriate escalation of 'outbreak' management. This kind of 'de-compromising' and 'de-escalating'is one of the major reasons why local hospitals and their laboratories perform strain typing for outbreak analysis. Thus, whether in a clinical or public health setting, the discriminatory or resolving power of a given epidemiological analysis is not solely dependent on a method or a method-pathogen combination but may be also be influenced by the pathogens' diversity (i.e. the more or less frequent appearance/epidemicity or endemicity of a specific type).

Choosing the 'best' method for typing

Whether considering strain typing from the clinical or public health perspective, the logical question is: what is the best method procedurally to use? However, there are a number of reasons why a 'one size fits all' answer to this question is impractical.

Considering first the clinical environment, as noted earlier, strain typing is commonly of value in assessing therapeutic concerns such as multisite infection or emergence of antimicrobial resistance in the individual patient, and transmission of problem pathogens within a limited patient population (e.g. a healthcare or family unit). In this context the key issues include: (i) having the required technical expertise, (ii) potential for automation/routine applicability, (iii) cost, (iv) required time-to-answer, (v) equipment maintenance and footprint size, (vi) intuitive data output and objective, standardisable, or automated interpretation, (vii) relevance of the typing result for further investigations (e.g. screening of staff) or for reporting to public health authorities.

It is logical to aspire to the most recently published cutting-edge method. However, the newest iteration of the most sophisticated and advanced technology is of little value if one does not have physical room for it, cannot afford it, properly operate it, or readily achieve clinically or epidemiologically relevant outcomes from the data generated. While one would never recommend gravitating to the lowest technological denominator for strain typing, to a large extent the 'best' method in a given clinical environment depends on the available resources addressing the issues noted above. In this context, as stated earlier, it is important to recognise that, regardless of sophistication, molecular strain typing commonly operates from an incomplete data set since all relevant clinical isolates may not be available and all isolate characteristics may not have been analysed, although the latter issue will be less of a concern in the future as WGS becomes more refined and widespread. In addition, communication between appropriate clinical interests (e.g. physician, laboratory, nursing, infection control) is vital to putting the 'incomplete'typing data into the fullest context for a meaningful outcome in terms of infection prevention and control.

Taken together, in addition to routine and real time strain typing, key elements for successful strain typing in the clinical setting most certainly include [3,15]: (i) initiation of strain typing by the hospital epidemiologist in consultation with infection control, infectious disease, and microbiology personnel, (ii) targeting of strain typing to investigate specific infectious disease issues such as an unusual increase in the rate of isolation of a pathogen, a cluster of infections in a particular healthcare unit, and multiple isolates with unusual (e.g. antibiotic susceptibility) characteristics, (iii) understanding that strain typing in the absence of epidemiological context and follow-up is an inefficient use of laboratory resources. Strain typing should supplement, not replace, careful epidemiological investigation.

To a large extent, the issues affecting approaches to strain typing for public health purposes are similar to those previously noted for local clinical efforts. However, there are important differences. The concerns of public health, while clinical in nature, are much broader in scope especially focusing on the transmission of problem pathogens on a local, regional, national, and international scale. Therefore, while financial and technical resources are generally more abundant at the regional/national level, the complexity of the necessary outcomes is greater as well. Effective communication to insure that the typing method's results are comparable between all laboratories involved is at the heart of a proper large-scale understanding of infectious disease occurrence and transmission. Everything from choice of typing method to data output and interpretation revolves around this issue. Thus, from a methodological standpoint the strain typing approach should: (i) be as standardised

as possible to be performed with similar efficiency, accuracy, and reproducibility in different participating laboratories, (ii) generate output that can be efficiently databased and shared, with interpretative criteria as objective as possible and a common terminology for strain type and subtype designations.

In this regard, sequence-based approaches hold the greatest promise. For example, SLST of the staphylococcal protein A gene (spa-typing) is effectively used in the epidemiological monitoring of specific *Staphylococcus aureus* strains (i.e. SeqNet; www.seqnet.org) with 540 laboratories from 51 countries submitting strains from 90 countries worldwide using the Ridom spa server as a common platform [16]. As noted earlier, approaches to WGS are rapidly being developed and refined with the potential to ultimately provide strain typing data ranging from key gene subsets [17] to total chromosomal comparison [18]. However, the success of the PulseNet System, designed by the United States Centers for Disease Control to investigate food-borne outbreaks [19], as well as refinements in VNTR-based analysis of pathogens such as meticillin-resistant *S. aureus* [5,20], illustrate that older molecular typing approaches also have potential for effective public health application.

Clinical and public health strain typing in perspective

Whether performed in a local clinical or more regional/national public health setting, the effective use of strain typing requires an understanding of both the pitfalls and the promises of the process. While the pitfalls can certainly be methodological, perhaps the most fundamental caveat, as noted above, is that strain typing is not a standalone method. Therefore, more information and communication is better than less. The scenario is not unlike an unfolding mystery story where one needs as much evidence as possible to figure out who 'did it.' For both local and larger-scale regional settings, the promise is a better understanding of the dynamics of infectious disease transmission with the hope of effective intervention (prevention, infection control, and treatment). Remarkable possibilities are on the horizon when one considers advances in genomic characterisation and the power of the Internet to facilitate the linking of strain typing analysis and databasing to other previously disparate data such as antimicrobial resistance (e.g. European Antimicrobial Resistance Surveillance Network (EARS-Net); www.ecdc.europa.eu/en/activities/surveillance/EARS-Net/Pages/index.aspx) and geographic information systems (GIS) as elegantly shown by the European Staphylococcal Reference Laboratory (SRL) working group (www.spatialepidemiology.net/srl-maps)[21] EpiScanGIS (www.episcangis.org), Global Network for Geospatial Health (GnosisGIS) (www.gnosisgis.org), and the World Health Organization (WHO)'s Public Health Mapping GIS effort (www.who.int/health_mapping/en). Most recently, during the *Escherichia coli* O104:H4 outbreak in Germany, open-source genomic analysis, available hardware/software resources and international expertise contributed tremendously to the rapid understanding of the pathogens' evolution, dissemination, and pathology [22]. Thus, for the future, the promises outweigh the pitfalls as molecular strain typing seeks to address enduring infectious disease issues with important morbidity, mortality, economic, and general quality of life implications.

References

1. Goering RV. Molecular typing techniques: state of the art. In: Tang YW, Stratton CW, editors. Advanced techniques in diagnostic microbiology. 2nd ed. New York (NY): Springer; 2013. p. 239-61.
2. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18(4):pii=20380. Available from: www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20380
3. 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.
4. Robinson DA, Enright MC. Evolutionary models of the emergence of meticillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother. 2003;47(12):3926-34.
5. Schouls LM, Spalburg EC, van LM, Huijsdens XW, Pluister GN, Van Santen-Verheuvel 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.
6. Goering RV, Ribot EM, Gerner-Smidt P. Pulsed-field gel electrophoresis: laboratory and epidemiologic considerations for interpretation of data. In: Persing DH, Tenover FC, Tang YW, Nolte FS, Hayden RT, Belkum A, et al., editors. Molecular microbiology. 2nd ed. Washington (DC): ASM Press; 2011. p. 167-77.
7. Higuchi R, Glyyensten U, Persing DH. Next-generation DNA sequencing and microbiology. In: Persing DH, Tenover FC, Tang YW, Nolte FS, Hayden RT, Belkum A, et al., editors. Molecular microbiology: diagnostic principles and practice. 2nd ed. Washington (DC): ASM Press; 2011. p. 301-12.
8. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
9. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 2012;8(8):e1002824.
10. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
11. Schatz MC, Langmead B, Salzberg SL. Cloud computing and the DNA data race. Nat Biotechnol. 2010;28(7):691-3.
12. Angiuoli SV, White JR, Matalka M, White O, Fricke WF. Resources and costs for microbial sequence analysis evaluated using virtual machines and cloud computing. PLoS One. 2011;6(10):e26624.
13. 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.
14. Hallin M, Deplano A, Roisin S, Boyart V, De Ryck R, Nonhoff C, et al. Pseudo-outbreak of extremely drug-resistant Pseudomonas aeruginosa urinary tract infections due to contamination of an automated urine analyzer. J Clin Microbiol. 2012;50(3):580-2.
15. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Infect Control Hosp Epidemiol. 1997;18(6):426-39.
16. Harmsen D, Claus H, Witte W, Rothganger J, Claus H, Turnwald D, et al. Typing of meticillin-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.
17. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012;158(Pt 4):1005-15.
18. Vogel U, Szczepanowski R, Claus H, Junemann S, Prior K, Harmsen D. Ion torrent personal genome machine sequencing for genomic typing of Neisseria meningitidis for rapid determination of multiple layers of typing information. J Clin Microbiol. 2012;50(6):1889-94.
19. 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.
20. 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 meticillin-resistant Staphylococcus aureus. J Clin Microbiol. 2012;50(7):2255-62.
21. 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.
22. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N Engl J Med. 2011;365(8):718-24.

The need for ethical reflection on the use of molecular microbial characterisation in outbreak management

B Rump (BRump@ggdmn.nl)[1], C Cornelis[2], F Woonink[1], M Verweij[2,3]
1. Municipal Health Service (GGD) Midden-Nederland, Zeist, the Netherlands
2. Department of Philosophy, Utrecht University, Utrecht, the Netherlands
3. Ethics Institute, Utrecht University, Utrecht, the Netherlands

Citation style for this article:
Rump B, Cornelis C, Woonink F, Verweij M. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. Euro Surveill. 2013;18(4):pii=20384. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20384

Article submitted on 20 June 2012 / published on 24 January 2013

Current thinking on the development of molecular microbial characterisation techniques in public health focuses mainly on operational issues that need to be resolved before incorporation into daily practice can take place. Notwithstanding the importance of these operational challenges, it is also essential to formulate conditions under which such microbial characterisation methods can be used from an ethical perspective. The potential ability of molecular techniques to show relational patterns between individuals with more certainty brings a new sense of urgency to already difficult ethical issues associated with privacy, consent and a moral obligation to avoid spreading a disease. It is therefore important that professionals reflect on the ethical implications of using these techniques in outbreak management, in order to be able to formulate the conditions under which they may be applied in public health practice.

Introduction

Recent advances in molecular microbial characterisation open up new scientific opportunities for a better understanding of not only the pathogenicity, evolution and spread of human pathogens, but also the epidemiology of the diseases they cause. Such progress has promising prospects for infectious disease control, particularly for real-time source and contact tracing in outbreak management [1]. Current thinking on the development of molecular microbiology characterisation techniques in public health focuses predominantly on the operational issues that need to be resolved [1]. The recommendations of the November 2011 expert consultation *Breakthroughs in molecular epidemiology of human pathogens - how to translate breakthroughs into public health practice*, organised by the European Centre for Disease Prevention and Control (ECDC), clearly outline the scientific hurdles that need to be overcome in order for public health to benefit from the recent scientific and technological advances in the rapidly evolving next generation sequencing technologies [1]. Despite the importance of these operational challenges, it is also essential to address the ethical difficulties associated with the use of microbial characterisation techniques in public health. The need for ethical guidance concerning the use of molecular typing methods is not new. Ethical challenges stemming from the introduction of molecular genomics have often been addressed in the context of population-level genomics and biobanking; such issues include those related to persons' autonomy and the patient–healthcare professional relationship. The use of these techniques in infectious disease control now raises similar ethical issues, in which individual interests and individual needs must be weighed against those of the public at large [2]. Due to recent scientific and technological advances in molecular microbial characterisation, the need for ethical guidance has now gained a new sense of urgency [3].

Although microbial characterisation techniques have primarily (and successfully) been used to benefit the general public's health, the results can also be used for other purposes, notably in support of legal or moral claims about responsibility and liability. For example, in 2007, in the Netherlands, genetic sequence analysis of HIV strains was used in a criminal case, in which the plaintiffs, who were recently infected with HIV, accused the defendants of deliberately administering them a subcutaneous injection of HIV-positive blood. The comparison of the genetic sequence analysis of the HIV strains of both parties was used as legal evidence [4].

Another example is a listeriosis outbreak linked to a food facility near Toronto, Canada, in 2008 [2]. Public health officials initially responded to the outbreak using traditional contact tracing and outbreak investigation. The food facility issued a voluntary recall of cold meat products before a confirmed linkage was available. Typing was used retrospectively to reduce uncertainty about the link between the 100 confirmed cases (23 deaths) and exposure to contaminated food from the facility. The resulting class-action lawsuits filed in four Canadian provinces were settled in December 2008 for US$ 27 million.

In these examples, the approach was successful from a medical perspective; however, from an ethical perspective, information that was initially collected for the purpose of public health was then also used in a legal context. Thus the results of genetic sequencing of infectious agents for purposes that go beyond protection of public health can yield important societal benefits, but can also facilitate legal claims (and hence economic risk) for individual persons or companies. And even in cases where individual persons or companies could do little to prevent an outbreak or avoid being a causal factor in transmitting disease, public perception of responsibility for infection may easily lead to stigmatisation and thus negatively affect the lives of the persons involved.

In this context, we highlight the most dominant ethical issues in regard to the use of molecular techniques. This is to facilitate further ethical reflection by public health professionals regarding the use of molecular techniques. We use the term to refer to a range of molecular microbial characterisation techniques that enable the linking of pathogens and that are now becoming increasingly available for real-time source and contact tracing.

Relational patterns between pathogens and people: a sense of urgency to the existing ethical debate

The ethical challenges associated with molecular techniques are mostly linked to their ability to give more precise information on the relational patterns between different microbes found in an outbreak [5-7]. Although the results of such techniques must be understood in the context of traditional epidemiological information – and even then, the most probable transmission route is rarely the only one possible – molecular techniques can allow more certainty on the relational patterns between microbes found in an outbreak. This inference about the directionality of transmission, may however, also specify the relational patterns between the people hosting them. This may be perceived by the public as an answer to the 'who infected whom?' question in an outbreak. While the ethical issues related to this question are not new, molecular techniques may heighten the level of certainty regarding such patterns and in this way introduce a sense of urgency to the ethical debate [8,9].

Moral obligation to avoid spreading a disease

That advanced sequencing technologies show potential relational patterns between people may fuel public discussions about who is responsible for infection or outbreaks. This is a complex issue with no simple conclusions; however, it is tempting to jump from information about 'who infected whom' to judgments about responsibility for infection. Attribution of responsibility to individuals for outbreaks of infectious diseases, however, is ethically problematic, even with the most sophisticated microbial molecular typing techniques. This is because although molecular microbial typing methods can help to elucidate potential transmission pathways, additional conditions are required before moral responsibility for the spread of infection can be attributed to individuals. More advanced molecular technology (in combination with epidemiological information) may be able to visualise certain transmission patterns in an outbreak, but this does not necessarily lead to factual conclusions on the cause of disease. Transmission of a microbe, for instance, may lead to colonisation, but colonisation may not necessarily lead to infection or subsequent disease. But even if we assume that transmission leads to disease, this does not make the source or actor morally responsible. The conditions for attributing moral responsibility for spreading a disease include numerous factors that need to be taken into account, for instance, knowledge of the risk, of the transmission pathways and ways to avoid infection, as well as competence to take adequate precautions [10]. Moreover, whether one can rightfully attribute moral responsibility will depend on whether it is reasonable to expect people to take precautions against infecting others and whether the infected persons could have easily protected themselves. Hence, judgments about moral responsibility are complex: even though molecular typing technologies may show relatively clear transmission pathways, this should not be considered as a sufficient basis for judgments about responsibility for infection. This is not to say that the laws of some countries may address this moral obligation to avoid spreading a disease and have specified what action is legally prohibited, required or permissible, attaching legal consequences for those who fail to act in line with such dictates.

Ownership of pathogens

In addition to this concept of a moral responsibility for infection, molecular techniques also place the concept of privacy in a new perspective. The question of privacy is associated in a way with the question of ownership. In bioethics, there already is a debate on who owns a biological specimen isolated from an individual at a certain moment in time [11], regarding whether a biological specimen (such as tissue, blood or stool) and the pathogen found in this specimen, in some way 'belong' to the individual they came from. In outbreak management, this question is further complicated by the fact that a number of pathogens are transmissable from person to person, which means that they may be seen as 'owned' by various persons over time.

Informed consent

Irrespective of the outcome of this ownership debate, privacy from a perspective of *ethical* and *legal* issues surrounding informed consent also need to be addressed when molecular techniques are used in outbreak management. There are various ethical and legal theories or accounts given of what informed consent exactly means and how it should be conducted in practice. From an ethical perspective, informed consent is concerned with the consent being 'informed',

'voluntary', and 'decisionally-capacitated', meaning that all information needs to be disclosed to a competent ('capacitated') patient, who understands all that has been disclosed, and that this patient voluntarily consents to treatment (or to a research subject when it comes to participation in research) [12]. This raises important questions about how these informed consent requirements could be conceptualised when using molecular techniques in outbreak management. One such question pertains to formulating information disclosure requirements: what (type and how much) information ought to be disclosed and comprehended in order for someone to be able to legitimately consent to any type of intervention or procedure proposed by a public health official? Intertwined with this is the question of who consent must be obtained from. Due to the fact that many individuals may be involved in an outbreak, and because sequence information about the pathogen in a particular infected individual may give rise to new information about, for example, relational patterns to other infected persons, the question of *who exactly*, of all the persons involved in an outbreak, should be consenting to the use of such technologies remains a pertinent one. Such information could be relevant to a number of parties involved in an outbreak for different reasons, and the interests of those parties in that information could, moreover, conflict with each other. Furthermore, informing all the parties may be seen as an unrealistic task, depending on the type and amount of information that needs to be disclosed and who must be informed. This is also relevant to the current management of outbreaks, but molecular techniques give more specificity about the directionality of transmission and can be used on a pathogen obtained from one person and interpreted along with information obtained from another person. This makes answering the question to whom disclosures should be made, who should agree to participate and whether full comprehension of the information in itself can be reached even more complex.

Return of results

Another issue that needs to be addressed when using molecular techniques in outbreak management is the concept of a 'return of results' duty. This concept pertains to the problem of how and to what extent, or whether (research) information needs to be returned to certain parties, for instance, the individual and/or the public. This is an issue well addressed in biobanking, where the debate focuses on treatment options or financial gain [11]. When it comes to outbreak management, however, the issue is more complex: here it is not only about the (financial or medical) interests of specific individuals directly associated with the intervention but also about the many parties involved in an outbreak. The interests and needs of specific individuals need to balanced with those of the general population. Furthermore, disclosure of information may be of immediate public health interest and, at the same time, be harmful to the people directly involved.

Legal perspective

A legal norm or duty and its justification are not the same as a moral norm or duty and its justification. Although the presence and adoption of legal duties are frequently justified (usually at least in part) by ethical arguments, what ultimately validates a legal norm is its recognition by a political and/or legal institution or authority. That is, a legal norm is operationalised through institutional rules and governance structures (ranging from laws and regulations to policies and guidelines). The law attempts to find a coherent position in balancing population interests versus individual freedoms [13]. The introduction of novel technologies into health systems often brings forth new ethical arguments and this may change the perspective on these population interests or individual freedoms. However the present legal norm cannot easily be changed and cannot even always be directly met by new jurisdiction [13].

When it comes to the legal framework for controlling infectious diseases and the protection of public health; using molecular techniques may not even be a problem in many European countries [14]. Public health law in many countries already makes surveillance legally possible without explicit patient consent [14]; however, to what extent this includes a legal possibility for microbiological research and molecular typing in outbreak management is not well defined.

Conclusion

In light of the ability of molecular techniques to show potential relational patterns between people and that this may fuel public discussions about who is responsible for an infection or outbreaks, it is essential to not only address operational challenges related to use of such techniques in outbreak management, but also to shape the conditions under which they can be used in practice. Reflection on these conditions may not result in closure of the ethical debate on topics such as privacy, consent and moral obligation to avoid infecting others, but it can offer guidance to public health professionals who use these techniques in source and contact tracing.

Call for ethical reflection

In this context, the Dutch Municipal Health Service GGD Midden-Nederland focuses on the ethical questions concerning the use of molecular typing techniques in the control of infectious diseases. Our current project, supported by the Dutch National Institute for Public Health and the Environment (RIVM) through the regional support fund for reinforcement of infectious disease control, aims at combining public health ethics with practice. We warmly invite public health professionals, especially microbiologists, to put their reflections on the conditions under which molecular techniques should be used in source and contact tracing in writing (send them by email to ethiektraining@ggdmn.nl before 15 March 2013).

Acknowledgments

This work is funded by the Dutch Ministry of Health Welfare and Sport, RIVM, through the regional support fund for reinforcement of infectious disease control.

References

1. Palm D, Johansson K, Ozin A, Friedrich AW, Grundmann H, Larsson JT, et al. Molecular epidemiology of human pathogens: how to translate breakthroughs into public health practice. Stockholm, November 2011. Euro Surveill. 2012;17(2):pii=20054. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20054
2. Fanoy E, De Neeling A. Molecular typing: use with care. Public Health Ethics. 2012;5(3):313-4.
3. Rump B, Woonink F. Ethical questions concerning the use of molecular typing techniques in the control of infectious diseases. Public Health Ethics. 2012;5(3):311-3.
4. van der Kuyl AC, Jurriaans S, Back NK, Sprenger HG, van der Werf TS, Zorgdrager F, et al. Unusual cluster of HIV type 1 dual infections in Groningen, the Netherlands. AIDS Res Hum Retroviruses. 2011;27(4):429-33.
5. van Belkum A. Molecular typing of micro-organisms: at the centre of diagnostics, genomics and pathogenesis of infectious diseases? J Med Microbiol. 2002;51(1):7-10.
6. Kretzschmar M, Gomes MG, Coutinho RA, Koopman JS. Unlocking pathogen genotyping information for public health by mathematical modeling. Trends Microbiol.2010;18(9):406-12.
7. Patel SJ, Graham PL 3rd. Use of molecular typing in infection control. Pediatr Infect Dis J. 2007;26(6):527-9.
8. Harris J, Holm S. Is there a moral obligation not to infect others? BMJ. 1995;311(7014):1215–7.
9. Verweij M. Obligatory precautions against infection. Bioethics. 2005;19(4):323-35.
10. Millar M. Moral permissibility and responsibility for infection. Public Health Ethics. 2012;5(3):314-7.
11. Hawkins AK, O'Doherty KC. "Who owns your poop?": insights regarding the intersection of human microbiome research and the ELSI aspects of biobanking and related studies. BMC Med Genomics. 2011;4:72.
12. Informed consent. In: Stanford Encyclopaedia of Philosophy, Stanford, CA: Stanford University. [Accessed 12 Dec 2012]. Available from: http://plato.stanford.edu/entries/informed-consent/
13. Bubela T, Yanow S. Molecular typing technology: a legal perspective. Public Health Ethics. 2012;5(3):317-20.
14. Lee LM, Heilig CM, White A. Ethical Justification for Conducting Public Health Surveillance Without Patient Consent. Am J Public Health. 2012;102(1):38-44.

Molecular-based surveillance of campylobacteriosis in New Zealand – from source attribution to genomic epidemiology

P Muellner (petra@epi-interactive.com)[1,2], E Pleydell[2], R Pirie[3], M G Baker[4], D Campbell[5], P E Carter[3], N P French[2]
1. Epi-interactive, Miramar, Wellington, New Zealand
2. mEpiLab, Institute of Animal, Biomedical and Veterinary Science, Massey University, Palmerston North, New Zealand
3. Institute of Environmental Science and Research, Kenepuru Science Centre, Porirua, New Zealand
4. Department of Public Health, University of Otago, Wellington, New Zealand
5. Ministry for Primary Industries, Wellington, New Zealand

Citation style for this article:
Muellner P, Pleydell E, Pirie R, Baker MG, Campbell D, Carter PE, French NP. Molecular-based surveillance of campylobacteriosis in New Zealand – from source attribution to genomic epidemiology. Euro Surveill. 2013;18(3):pii=20365. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20365

Article submitted on 27 June 2012 / published on 17 January 2013

Molecular-based surveillance of campylobacteriosis in New Zealand contributed to the implementation of interventions that led to a 50% reduction in notified and hospitalised cases of the country's most important zoonosis. From a pre-intervention high of 384 per 100,000 population in 2006, incidence dropped by 50% in 2008; a reduction that has been sustained since. This article illustrates many aspects of the successful use of molecular-based surveillance, including the distinction between control-focused and strategy-focused surveillance and advances in source attribution. We discuss how microbial genetic data can enhance the understanding of epidemiological explanatory and response variables and thereby enrich the epidemiological analysis. Sequence data can be fitted to evolutionary and epidemiological models to gain new insights into pathogen evolution, the nature of associations between strains of pathogens and host species, and aspects of between-host transmission. With the advent of newer sequencing technologies and the availability of rapid, high-coverage genome sequence data, such techniques may be extended and refined within the emerging discipline of genomic epidemiology. The aim of this article is to summarise the experience gained in New Zealand with molecular-based surveillance of campylobacteriosis and to discuss how this experience could be used to further advance the use of molecular tools in surveillance.

Controlling campylobacteriosis – recent successes

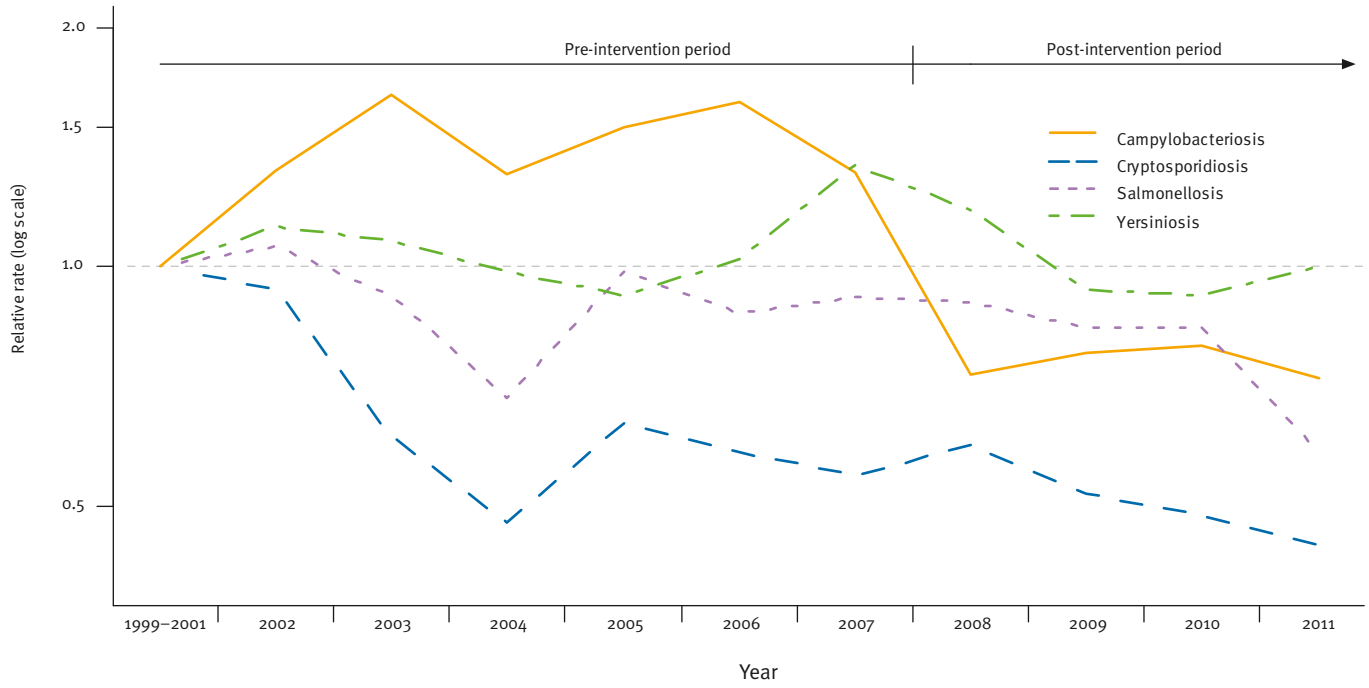
Molecular tools are being used increasingly to inform the control of enteric zoonosis worldwide [1] and to meet a wide range of public health aims and objectives [2-4]. In New Zealand, a country with a historically high rate of campylobacteriosis notifications [5,6], results from molecular-based surveillance in a sentinel site founded in 2005 – where human cases and potential sources were sampled and typed by multilocus sequence typing (MLST) simultaneously over consecutive years [7,8] – provided strong evidence that a large proportion of human cases were linked poultry meat consumption. These findings contributed to a mounting body of evidence [5,9] and stimulated the implementation of regulatory and voluntary control strategies along the poultry supply chain. They were announced in 2007 and fully implemented in 2008 (when they became mandatory) [10], resulting in a 50% reduction in disease incidence of cases in 2008 compared with the previous high level during 2002 to 2006 [10,11].

Campylobacteriosis notifications in humans were markedly above the reference value until 2008, when the incidence dropped considerably (Figure 1); a likely effect of a reduction in poultry-associated cases due to the implementation of the control strategies in the poultry supply chain [10,11]. No comparable changes in the annual incidence of other enteric notifiable diseases were observed over the same time period (2002–2011) (Figure 1). Sustained decline in campylobacteriosis case numbers has been shown to have additional health and economic benefits by, for example reducing the incidence of Guillain–Barré syndrome, an autoimmune condition associated with prior *Campylobacter* spp. infection [12].

Furthermore, in the New Zealand sentinel surveillance site, a dominant poultry-associated MLST sequence type of *C. jejuni* (ST-474) was identified that, to date, has been reported rarely from other countries. Before the implementation of the poultry interventions, ST-474 accounted for 30% of human cases in the sentinel site [14,15], but in 2010–11, it was isolated from less than 5% of cases [16]. Figure 2 shows the dramatic reduction in two major poultry-associated genotypes, ST-474 and ST-48 (Figure 2, panel A), and provides a comparison with other STs over the same time period (Figure 2, panels B and C).

Figure 1

Relative rates[a] of notification of campylobacteriosis, cryptosporidiosis, salmonellosis and yersiniosis, New Zealand, 2002–2011 compared with 1999–2001



[a] Rates were calculated using a negative binomial model, which was used to estimate the change in incidence between each year from 2002 to 2011 and the reference period of interest, 1999–2001, as previously described by Henao et al. [13]. Values above the reference line indicate increases in notification incidence and points below the line show decreases, relative to the 1999–2001 reference period.

The pre- and post-intervention periods refer to the implementation of a number of control measures in the poultry supply chain by the regulatory authority. The annual incidence of other enteric notifiable diseases (cryptosporidiosis, salmonellosis and yersiniosis) over the same time period is displayed to show that notification rates were stable for other comparable disease and that the drop in campylobacteriosis notifications was not a surveillance artefact.

Focused molecular epidemiological studies have been contributing to our understanding of the epidemiology of this widespread disease both in New Zealand and elsewhere [7,14,15,17]. For example, the association between ruminant-associated genotypes and preschool-age children (0–5 years of age) in rural areas has provided evidence for direct contact with faecal material being the foremost infection route in this high-incidence group [14].

This is of high relevance for the development and evaluation of appropriate, country-specific control strategies to decrease the human disease burden. Since the number of human cases linked to poultry has fallen in New Zealand, there has been a relative increase in importance of ruminant strains of *C. jejuni*, and ongoing work is investigating the complex epidemiology of *Campylobacter* in ruminant [18] and wildlife sources [19]. While this article describes the MLST-supported *Campylobacter* surveillance conducted at the sentinel site, other typing approaches are used to increase resolution of the molecular analysis. For example, research is currently underway to further differentiate between exposure to ruminant-associated *Campylobacter* subtypes of food and non-food origin to refine attribution estimates using antigen gene sequence typing [20], ribosomal MLST [21] and targeted genes identified by whole genome analysis [22]. In this article, we summarise the experience gained in New Zealand and discuss how this experience could be used to further advance the use of molecular tools in surveillance.

What have we learned?

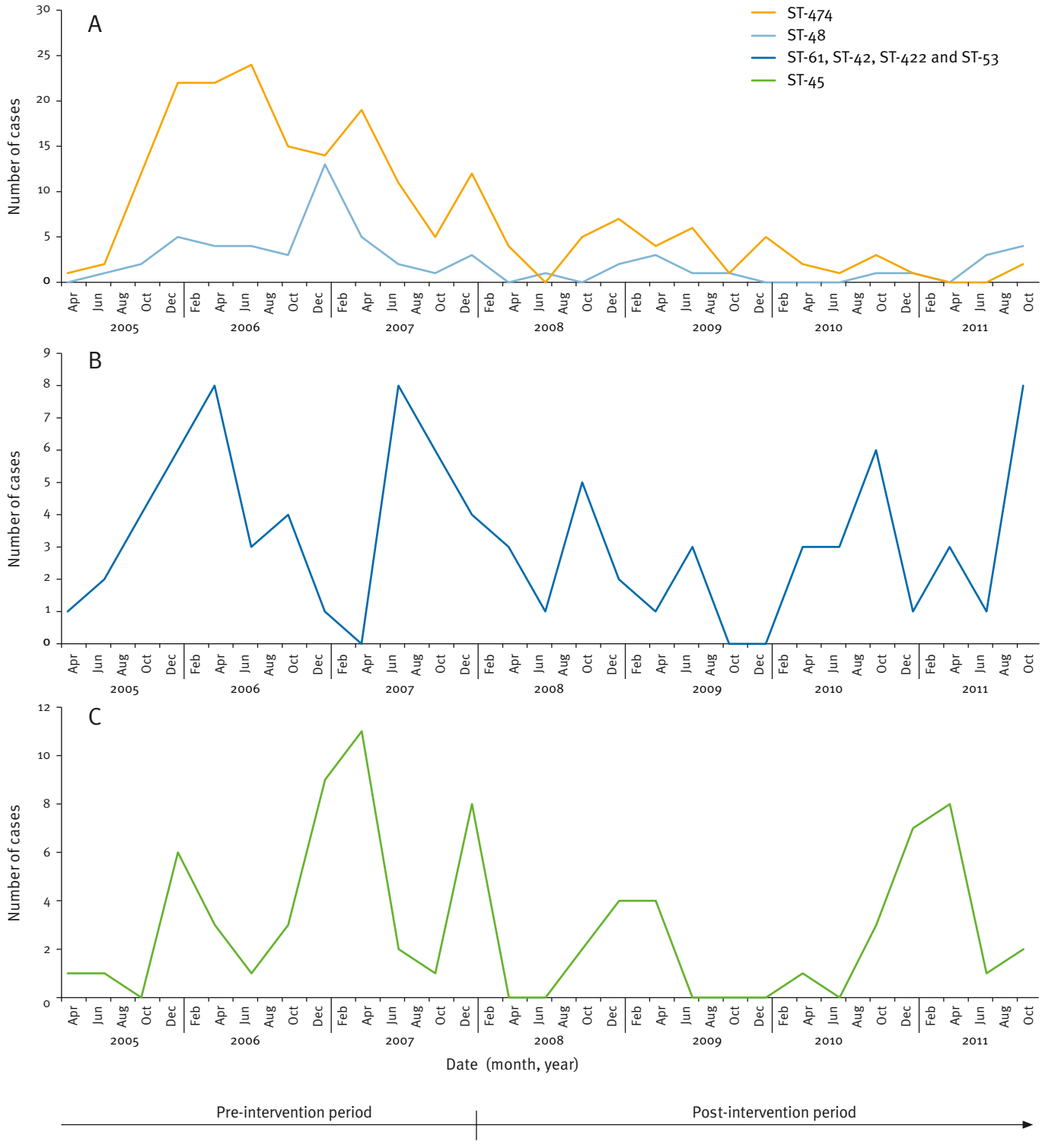
Experience from New Zealand and elsewhere has provided insight into key aspects of molecular-based surveillance. These include the following: (i) its application to both control-focused and strategy-focussed surveillance; (ii) a change in our definition of epidemiological response variables; and (iii) the emergence of genomic epidemiology.

Application of molecular tools to disease surveillance

The framework developed by Baker et al. [23], which differentiates between control-focused and strategy-focused surveillance, provides a meaningful way to

Figure 2

Human cases of campylobacteriosis caused by poultry- and ruminant-associated *Campylobacter jejuni* MLST types, as well as a ubiquitous ST in a sentinel surveillance site, New Zealand, 2005–2011



MLST: multilocus sequence typing; ST: sequence type.

Panel A shows the time series of human campylobacteriosis cases with two poultry-associated genotypes, ST-474 and ST-48 and illustrates the drop in the number of cases following interventions in the poultry production chain.

Panrel B shows the trend in human campylobacteriosis cases with ruminant-associated genotypes ST-61, ST-42, ST-422 and ST-53.

Panel C shows the time series of human campylobacteriosis cases with the ubiquitous ST-45.

categorise molecular approaches to disease surveillance. Approaches that are suitable for control-focused surveillance, such as those used in an outbreak setting, are potentially of lesser value for strategy-focused surveillance, where the aim is often to monitor long-term changes in epidemiology [24,25], and vice versa.

The purpose of control-focused surveillance is 'to identify each occurrence of a particular disease, hazard, or other health-related event that requires a specific response, and to support the delivery of an effective intervention' [23]. Such surveillance requires methods that have a high degree of timeliness, sensitivity and security (i.e. that can be maintained on an ongoing basis) [23]. The molecular typing tools and associated modelling approaches required for this objective need to be capable of identifying genotypes that indicate a common source of disease or highlight a particular transmission pathway. Often, but not always, this is achieved through highly discriminatory typing tools.

Using a recently developed model-based tool for identifying clusters of campylobacteriosis cases related in space and time [26], eight cases in a small area of New Zealand's North Island were identified as having a high probability (>0.8) of being part of an anomalous cluster (i.e. they were more spatially and temporally localised than would be expected given the average temporal and spatial patterns). Inspection of the epidemiological information linked to each case revealed that they were reported within a single two-week period and typing data showed that they were all the same MLST sequence type (ST-520). When compared with a larger database of over 3,000 sequence types isolated from humans, animals and food in New Zealand it was shown that this type was associated with ruminants in New Zealand, but was a relatively uncommon cause of human infection. This finding triggered a more detailed investigation into the cases, requiring further contact with some affected individuals, which revealed that all cases had consumed unpasteurised (raw) milk – a relatively rare risk factor – and that 7/8 cases reported purchasing the milk from the same source farm. This combination of epidemiological information and typing data lead to a local response and also informed the ongoing debate on the national policy concerning the sale of raw milk.

The purpose of strategy-focused surveillance is 'to provide information to support prevention strategies to reduce population health risk, such as describing the epidemiology of the annual influenza season and the characteristics of the seasonal influenza viruses' [23]. Such surveillance requires methods that have a high degree of representativeness, completeness (referring to the data recorded with each event) and validity [23]. Different molecular and modelling approaches are required in this instance, with the optimal tools providing information on the long-term epidemiology of a pathogen rather than short-term changes. An example is the recent emergence of new approaches to source attribution using molecular subtyping, which has been used successfully in several countries to understand the relative contribution of different sources to the burden of human campylobacteriosis and salmonellosis [27-30]. Source attribution models based on microbial subtyping were initially developed in Denmark as a tool for salmonellosis risk management [31]; they provide estimates of the number of human cases originating from different sources or reservoirs based on a comparison of genotypes [31,32].

In New Zealand, attribution models were adapted to data from the MLST surveillance site. Two models were used, a population genetics-based attribution model [32] and the microbial subtyping-based model by Hald et al. [31], to quantify the contribution of selected sources to the human disease burden. These studies revealed that between 2005 and 2008, poultry was the leading source of human campylobacteriosis, causing an estimated 58–76%, of notified cases [8]. Contributions by individual poultry suppliers showed wide variation and supplier specific strains were detected [15]. The use of these models to monitor changes over time and to assess the effectiveness of interventions is ongoing [10,11].

Re-defining response and explanatory variables using molecular tools – a new epidemiological approach to inform surveillance?

A common starting point of epidemiology is seeking non-random associations between response variables and potential explanatory variables. Regression modelling, for example, may be used to identify statistically significant predictors of increased risk of adverse health effects [33]. However, the use of such traditional methods for quantifying the contribution of different sources of campylobacteriosis to the disease burden in New Zealand (notably case–control studies [9,34], which identified poultry as the major source of human infection) had not provided sufficient compelling evidence for decision-makers to invest in controlling the poultry source. The epidemiology of campylobacteriosis is challenging: as a multi-host pathogen, infection with *C. jejuni* is associated with a large number of risk factors [35] and human cases arising from exposure to different sources may have very different risk factors, some of which may even be protective for some sources and increase the risk for others.

Using molecular tools, pathogen genetics and evolution can be incorporated into epidemiological modelling to make inferences about disease or transmission risks rather than simply relying on the association of response and explanatory variables. Such tools can be used to refine outcome variables, for example by using case–case comparison of poultry- and ruminant-associated cases of campylobacteriosis to identify more subtle associations [14,17] or to investigate the cause of a disease outbreak [24]. However, greater epidemiological gains are likely to be made when models combine pathogen evolution and transmission in

an integrated way [36,37]. This may be best achieved by modelling a relatively low number of isolates with high-coverage sequence data, such as increasingly available full genome data [38] or a larger number of isolates with low coverage such as a 7-locus MLST scheme. The additional information provided even by routinely applied molecular tools such as pulsed-field gel electrophoresis (PFGE) adds to our understanding of epidemiological variables. For example, the level of similarity and relatedness of restriction-enzyme profiles in the analysis of a food-borne outbreak can be directly used to refine epidemiological investigations. It is the synergy between the epidemiological and typing information that makes molecular tools so powerful and novel modelling approaches are constantly being developed to advance research at this interface [39,40].

Into the future: genomic epidemiology

New modelling approaches are being adopted to utilise the abundance of molecular data available [24,39,40]. Bell et al. [41] argue that the enormous volumes of data that can be provided by new technology provides many challenges for data management and analysis, and that we have entered a new area of data-intensive science that requires specialised skills and analytical tools. This argument holds true in the area of molecular epidemiology: next generation high throughput sequencing has vastly increased the availability of pathogen genome sequence data [38] and as the costs decrease, these tools will be more frequently incorporated into epidemiological studies and surveillance. By fitting statistical genetics and epidemiological models to sequence data, and combining these within a single framework [42], new insights can be gained into pathogen evolution, the nature of associations between strains of pathogens and host species, the timing of emergence, origin and geographical spread of pathogens, and aspects of between-host transmission [43]. Furthermore, advances in statistical methods for modelling evolutionary ancestry are resulting in better reconstructions of pathogen genealogies and improved estimates of evolutionary parameters. Although complex in nature, these models can be extremely valuable – for example, they can be used to enable the contribution of different sources and transmission pathways to the human disease burden to be determined [32].

In New Zealand, whole genome sequencing is being used to understand the evolution of epidemiologically important strains of *C. jejuni* and identify potential markers for host association [44]. This may help to improve the discrimination of sources of human infection, such as between cattle and sheep, and result in more precise source attribution estimates. Similarly, full genome sequence data from multiple *Campylobacter* isolates and *Escherichia coli* O157 are being combined with phenotypic microarray data to improve the understanding of the relationship between phenotype and genotype. The identification of genetic markers for stress resistance, such as pH, temperature, oxidative stress, and freeze-thaw [45], could help to determine which sources and transmission pathways strains isolated from humans have been acquired from, further refining attribution studies and strategy-focused surveillance.

By furthering our understanding of host associations with particular strains of pathogens, and the relative rates of transmission between animals and humans, the melding of statistical genetics and epidemiology with partial and full genome sequence data will further inform and refine control strategies for enteric pathogens in New Zealand and elsewhere.

Conclusion

New Zealand provides a distinctive island ecosystem in which to study infectious diseases [46]. The relative isolation and management of farmed livestock has contributed to the epidemiology and population structure of microbial pathogens. For example, the country's poultry industry is structured in a way that is different to most countries, with no importation of untreated poultry products and freedom from several important poultry diseases such as Newcastle disease and *Salmonella enterica* serovar Enteritidis PT4. Furthermore, the production of poultry meat is highly integrated, with three companies supplying about 90% of all chicken meat [15]. In addition a risk management strategy developed by the regulator supports a strong collaboration with researchers and science-based decision-making [47]. While the situation in other countries is likely to be more complex, for example through the presence of federal regulations or the risks associated with poultry importations, lessons learned from New Zealand can be applied elsewhere.

The New Zealand approach, which includes the first large scale implementation of effective regulatory *Campylobacter* control measures in broilers, is of high relevance internationally, including Europe. Findings have been incorporated in scientific opinions of the European Food Safety Authority. In 2008, it was acknowledged that the MLST approach to source attribution developed in New Zealand may be the way forward [48] and the approach is being used in several European countries, including the Netherlands and Scotland [2,49]. The New Zealand experience was also included in an assessment of the extent to which meat derived from broilers contributes to human campylobacteriosis at the European Union level [50].

The molecular tools deployed in epidemiological and evolutionary analyses clearly need to be fit-for-purpose. Ideally, during their development phase, measures of their utility in specific settings, such as discriminatory power and the strength of association between genotype and host, should be considered and attempts made to optimise their performance for the outcome in mind. In the case of 7-locus MLST, for example, retrospective analyses have shown this to be a valuable approach for certain types of surveillance,

including reservoir attribution, but the method was not designed for this purpose and an alternative approach based on a different set of gene targets may perform better and be more cost-effective. Equally important are rigorous sampling size considerations and guidance on the number of samples from different sources to acquire a desired level of precision – for example, in source attribution estimates. Further work will be necessary to develop expert agreement and sound working principles on these matters.

The field of molecular epidemiology is continually evolving and its role in advancing our ability to understand and control infectious diseases will also keep increasing. Its interdisciplinary nature will provide key support to One Health approaches to disease control, by supplementing medical and veterinary expertise with an in-depth understanding of the molecular biology of pathogens. As genotyping approaches and analytical models continue to evolve, an understanding of the complex interface of both disciplines becomes a crucial element of molecular-based disease surveillance. In New Zealand, we have learned that close collaboration between laboratories and epidemiologists is extremely important for the success of molecular-based surveillance: in our example, this started when the sentinel surveillance site was first set up. In a small and geographically isolated country, such early collaboration is likely to be more easily achieved; nevertheless, the general principle still applies and could add value to molecular surveillance in other countries.

Acknowledgments

The authors are thankful to all members of the Hopkirk Molecular Epidemiology Team, Environmental Science and Research (ESR), MidCentral Health, Public Health Services, Science and Risk Assessment, Ministry for Primary Industries (MPI) and MedLab Central (in particular Lynn Rogers) for their contributions. This work was supported by MPI through their Postgraduate PhD Program 'Estimating the contribution of different food sources to the burden of human campylobacteriosis in the Manawatu' and the project 'Enhancing surveillance of potentially foodborne enteric diseases in New Zealand: human campylobacteriosis in the Manawatu' (FDI / 236 / 2005).

References

1. Foley SL, Lynne AM, Nayak R. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. Infect Genet Evol. 2009;9(4):430-40.
2. Sheppard SK, Dallas JF, Strachan NJ, MacRae M, McCarthy ND, Wilson DJ, et al. Campylobacter genotyping to determine the source of human infection. Clin Infect Dis. 2009;48(8):1072-8.
3. Swaminathan B, Gerner-Smidt P, Ng LK, Lukinmaa S, Kam KM, Rolando S, et al. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis. 2006;3(1):36-50.
4. Fournet N, Baas D, van Pelt W, Swaan C, Ober HJ, Isken L, et al. Another possible food-borne outbreak of hepatitis A in the Netherlands indicated by two closely related molecular sequences, July to October 2011. Euro Surveill. 2012;17(6):pii=20079. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20079
5. Baker M, Wilson N, Ikram R, Chambers S, Shoemack P, Cook G. Regulation of chicken contamination urgently needed to control New Zealand's serious campylobacteriosis epidemic. N Z Med J. 2006;119(1243):U2264.
6. Baker MG, Sneyd E, Wilson NA. Is the major increase in notified campylobacteriosis in New Zealand real? Epidemiol Infect. 2007;135(1):163-70.
7. McTavish SM, Pope CE, Nicol C, Sexton K, French N, Carter PE. Wide geographical distribution of internationally rare Campylobacter clones within New Zealand. Epidemiol Infect. 2008;136(9):1244-52.
8. Mullner P, Spencer SE, Wilson DJ, Jones G, Noble AD, Midwinter AC, et al. Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. Infect Genet Evol. 2009;9(6):1311-9.
9. Eberhart-Phillips J, Walker N, Garrett N, Bell D, Sinclair D, Rainger W, et al. Campylobacteriosis in New Zealand: results of a case-control study. J Epidemiol Community Health. 1997;51(6):686-91.
10. Muellner P, Marshall JC, Spencer SE, Noble AD, Shadbolt T, Collins-Emerson JM, et al. Utilizing a combination of molecular and spatial tools to assess the effect of a public health intervention. Prev Vet Med. 2011;102(3):242-53.
11. Sears A, Baker MG, Wilson N, Marshall J, Muellner P, Campbell DM, et al. Marked national decline in campylobacteriosis following interventions aimed at poultry. Emerg Infect Dis. 2011;17(6):1007-15.
12. Baker MG, Kvalsvig A, Zhang J, Lake RJ, Sears A, Wilson N. Declining Guillain-Barré syndrome after campylobacteriosis control, New Zealand, 1988-2010. Emerg Infect Dis. 2012;18(2):226-33.
13. Henao OL, Scallan E, Mahon B, Hoekstra RM. Methods for monitoring trends in the incidence of foodborne diseases: Foodborne Diseases Active Surveillance Network 1996-2008. Foodborne Pathog Dis. 2010;7(11):1421-6.
14. Mullner P, Shadbolt T, Collins-Emerson JM, Midwinter AC, Spencer SE, Marshall J, et al. Molecular and spatial epidemiology of human campylobacteriosis: source association and genotype-related risk factors. Epidemiol Infect. 2010;3(10):1372-83.
15. Müllner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SE, van derLogt P, et al. Molecular epidemiology of Campylobacter jejuni in a geographically isolated country with a uniquely structured poultry industry. Appl Environ Microbiol. 2010;76(7):2145-54.
16. French NP, Marshall J. Source attribution July 2010 to September 2011 of human Campylobacter jejuni cases from the Manawatu. Completion of sequence typing of human and poultry isolates and source attribution modelling. Palmerston North: Massey University; 2012.
17. Bessell PR, Rotariu O, Innocent GT, Smith-Palmer A, Strachan NJ, Forbes KJ, et al. Using sequence data to identify alternative routes and risk of infection: a case-study of campylobacter in Scotland. BMC Infect Dis. 2012;12:80.
18. French NP, Muellner P, Pleydell E. Attribution of potentially foodborne enteric diseases: human campylobacteriosis - ruminant attribution. Palmerston North: Massey University; 2010.
19. French NP, Midwinter A, Holland B, Collins-Emerson J, Pattison R, Colles F, et al. Molecular epidemiology of Campylobacter jejuni isolates from wild-bird faecal material in children's playgrounds. Appl Environ Microbiol. 2008;75(3):779-83.
20. Colles FM, Maiden MC. Campylobacter sequence typing databases: applications and future prospects. Microbiology. 2012;158(Pt 11):2695-709.

21. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012;158(Pt 4):1005-15.
22. Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, Kwan E, et al. Whole-genome comparison of two Campylobacter jejuni isolates of the same sequence type reveals multiple loci of different ancestral lineage. PLoS One. 2011;6(11): e27121.
23. Baker MG, Easther S, Wilson N. A surveillance sector review applied to infectious diseases at a country level. BMC Public Health. 2010;10:332.
24. Muellner P, Zadoks RN, Perez AM, Spencer SE, Schukken YH, French NP. The integration of molecular tools into veterinary and spatial epidemiology. Spat Spatiotemporal Epidemiol.. 2011;2(3):159-71.
25. Struelens MJ, De Gheldre Y, Deplano A. Comparative and library epidemiological typing systems: outbreak investigations versus surveillance systems. Infect Control Hosp Epidemiol. 19(8):565-9.
26. Spencer SE, Marshall J, Pirie R, Campbell D, French NP. The detection of spatially localised outbreaks in campylobacteriosis notification data. Spat Spatiotemporal Epidemiol. 2011;2(3):173-83.
27. Mullner P, Jones G, Noble A, Spencer SE, Hathaway S, French NP. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Anal. 2009;29(7):970-84.
28. Strachan NJ, Gormley FJ, Rotariu O, Ogden ID, Miller G, Dunn GM, et al. Attribution of Campylobacter infections in northeast Scotland to specific sources by use of multilocus sequence typing. J Infect Dis. 2009;199(8): 1205-8.
29. Mangen MJ, Batz MB, Käsbohrer A, Hald T, Morris JG Jr, Taylor M, et al. Integrated approaches for the public health prioritization of foodborne and zoonotic pathogens. Risk Anal. 2010;30(5):782-97.
30. Pires SM, Evers EG, van Pelt W, Ayers T, Scallan E, Angulo FJ, et al. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathog Dis. 2009;6(4):417-24.
31. Hald T, Vose D, Wegener HC, Koupeev T. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal. 2004;24(1):255-69.
32. Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E, et al. Tracing the source of campylobacteriosis. PLoS Genetics. 2008;4(9):e1000203.
33. Cox LA Jr. Detecting causal nonlinear exposure-response relations in epidemiological data. Dose Response. 2006;4(2):119-32.
34. Savill M, Hudson A, Devane M, Garrett N, Gilpin B, Ball A. Elucidation of potential transmission routes of Campylobacter in New Zealand. Water Sci Technol.. 2003;47(3):33-8.
35. Danis K, Di Renzi M, O'Neill W, Smyth B, McKeown P, Foley B, et al. Risk factors for sporadic Campylobacter infection: an all-Ireland case-control study. Euro Surveill. 2009 Feb 19;14(7):pii=19123. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=1912
36. Tibayrenc M. Bridging the gap between molecular epidemiologists and evolutionists. Trends Microbiol. 2005;13(12):575-80.
37. Archie EA, Luikart G, Ezenwa VO. Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol Evol. 2009;24(1):21-30.
38. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
39. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9):e1000520.
40. Matthews L, Reeve R, Woolhouse ME, Chase-Topping M, Mellor DJ, Pearce MC, et al. Exploiting strain diversity to expose transmission heterogeneities and predict the impact of targeting supershedding. Epidemics. 2009;1(4):221-9.
41. Bell G, Hey T, Szalay A. Computer science. Beyond the data deluge. Science. 2009;323(5919):1297-8.
42. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303(5656):327-32.
43. Kühnert D, Wu CH, Drummond AJ. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. Infect Genet Evol. 2011;11(8):1825-41.
44. Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, Kwan E, et al. Whole-genome comparison of two Campylobacter jejuni isolates of the same sequence type reveals multiple loci of different ancestral lineage. PLoS One. 2011;6(11):e27121.
45. Lee K, French NP, Jones G, Hara-Kudo Y, Iyoda S, Kobayashi H, et al. Variation in stress resistance patterns among stx genotypes and genetic lineages of Shiga toxin-producing Escherichia coli O157. Appl Environ Microbiol. 2012;78(9):3361-8.
46. Crump JA, Murdoch DR, Baker MG. Emerging infectious diseases in an island ecosystem: the New Zealand perspective. Emerg Infect Dis. 2001;7(5):767-72.
47. New Zealand Food Safety Authority (NZFSA). Campylobacter risk management strategy 2008-2011. Wellington: NZFSA; 2008. Available from: http://www.foodsafety.govt.nz/elibrary/industry/Campylobacter_Risk-Aims_Acheive.pdf
48. Scientific Opinion of the Panel on Biological Hazards on a request from EFSA on Overview of methods for source attribution for human illness from foodborne microbiological hazards. EFSA Journal. 2008;764:1- 43. Available from: http://www.efsa.europa.eu/it/scdocs/doc/764.pdf
49. Mughini Gras L, Smid JH, Wagenaar JA, de Boer AG, Havelaar AH, Friesema IH, et al. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One. 2012;7(8):e42599.
50. EFSA Panel on Biological Hazards (BIOHAZ). Scientific opinion on quantification of the risk posed by broiler meat to human campylobacteriosis in the EU. EFSA Journal. 2010;8(1):1437. Available from: http://www.efsa.europa.eu/en/efsajournal/doc/1437.pdf

A note from the editors: molecular epidemiology of human pathogens – current use and future prospects

Eurosurveillance editorial team (eurosurveillance@ecdc.europa.eu)[1]
1. European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden

Citation style for this article:
Eurosurveillance editorial team. A note from the editors: molecular epidemiology of human pathogens – current use and future prospects. Euro Surveill. 2013;18(3):pii=20374. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20374

Article published on 17 January 2013

While it is clear that surveillance and outbreak investigations are increasingly supported by advanced molecular approaches, it is less clear how the future of such tools will evolve. Even if they will most probably not render traditional epidemiological methods superfluous, they will certainly gain importance and a number of questions concerning their use in public health remain to be answered. How will both approaches interact in the future? Will they work hand in hand? How will the current operational constraints be overcome, to allow these advanced techniques to be used in public health practice?

In response to a call for papers [1], a special issue is being published. In the first part, we focus on examples of the value and opportunities of molecular methods in analysing a number of diseases. In the second part, we draw attention to issues related to their wider use in surveillance, prevention and control of infectious diseases. In so doing, we hope to stimulate discussion and add to the debate on the role and potential of modern molecular microbiology to inform public health action.

References

1. Call for papers for a special issue on molecular epidemiology of human pathogens. Euro Surveill. 2012;17(2):pii=20053. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20053

ECDC starts pilot phase for collection of molecular typing data

I van Walle (ivo.van.walle@ecdc.europa.eu)[1]
1. European Centre for Disease Prevention and Control, Stockholm, Sweden

Citation style for this article:
van Walle I. ECDC starts pilot phase for collection of molecular typing data. Euro Surveill. 2013;18(3):pii=20357. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20357

Article published on 17 January 2013

The European Surveillance System (TESSy) molecular surveillance service (MSS) was launched in late November 2012. This new service enables European Union and European Economic Area countries to upload detailed molecular typing data for *Salmonella*, *Listeria*, *E. coli* (VTEC/STEC) and multidrug-resistant *M. tuberculosis* (MDR-TB) to the MSS database, hosted by the European Centre for Disease Prevention and Control (ECDC). So far, 13 countries are providing data for the three food and waterborne pathogens, and it is likely that more countries will join the system in 2013. For *M. tuberculosis*, 23 countries are currently providing molecular typing data. It is likely that in the future, more pathogens will be added to the MSS.

The objective of the data is to (i) improve the speed of detection of dispersed international outbreaks, (ii) improve trace-back of the source of an outbreak and identify risk factors, and (iii) improve investigation of transmission chains across the EU and globally, and (iv) improve Member State response to outbreaks.

Nominated users in countries can query ECDC databases to see if isolates that are genetically similar to isolates found in their country have been detected in other countries. A team of curators guarantees the quality of the data and also performs routine cluster detection. If a cluster is found, countries can then use this cluster information for possible public health action. In such cases, the ECDC Epidemic Intelligence Information System will be used as before to help coordinate action across countries for *Salmonella*, *Listeria* and *E. coli* whereas the European Reference Laboratory Network for TB would be used in the case of MDR-TB. An evaluation of the pilot phase is planned for the end of 2013.

National Bulletins

Austria

Mitteilungen der Sanitätsverwaltung
Bundesministerium für Gesundheit Familie und Jugend, Vienna
Monthly, print only. In German.
http://www.bmgfj.gv.at/cms/site/thema.html?channel=CH0951

Belgium

Vlaams Infectieziektebulletin
Department of Infectious Diseases Control, Flanders
Quarterly, print and online. In Dutch, summaries in English.
http://www.infectieziektebulletin.be

Bulletin d'information de la section d'Epidémiologie
Institut Scientifique de la Santé Publique, Brussels
Monthly, online. In French.
http://www.iph.fgov.be/epidemio/epifr/episcoop/episcoop.htm

Bulgaria

Bulletin of the National Centre of Infectious and Parasitic Diseases, Sofia
Print version. In Bulgarian.
http://www.ncipd.org/

Cyprus

Newsletter of the Network for Surveillance and Control of Communicable Diseases in Cyprus
Medical and Public Health Services, Ministry of Health, Nicosia
Biannual, print and online. In Greek.
http://www.moh.gov.cy

Czech Republic

Zpravy CEM (Bulletin of the Centre of
Epidemiology and Microbiology)
Centrum Epidemiologie a Mikrobiologie Státního
Zdravotního Ústavu, Prague
Monthly, print and online. In Czech, titles in English.
http://www.szu.cz/cema/adefaultt.htm

EPIDAT (Notifications of infectious diseases in the Czech Republic)
http://www.szu.cz/cema/epidat/epidat.htm

Denmark

EPI-NEWS
Department of Epidemiology, Statens Serum Institut, Copenhagen
Weekly, print and online. In Danish and English.
http://www.ssi.dk

Finland

Kansanterveyslaitos
Department of Infectious Disease Epidemiology, National Public Health Institute, Helsinki
Monthly, print and online. In Finnish.
http://www.ktl.fi/portal/suomi/osastot/infe/tutkimus/tartuntatautien_seuranta/tartuntatautilaakarin_kommentit/

France

Bulletin épidémiologique hebdomadaire
Institut de veille sanitaire, Saint-Maurice Cedex
Weekly, print and online. In French.
http://www.invs.sante.fr/beh/default.htm

Germany

Epidemiologisches Bulletin
Robert Koch-Institut, Berlin
Weekly, print and online. In German.
http://www.rki.de/DE/Content/Infekt/EpidBull/epid__bull__node.html

Greece

HCDCP Newsletter
Hellenic Centre for Disease Control and Prevention (HCDCP/KEELPNO), Athens
Monthly, online. In English and Greek.
http://www2.keelpno.gr/blog/?lang=en

Hungary

Epinfo (az Országos Epidemiológiai Központ epidemiológiai információs hetilapja)
National Center For Epidemiology, Budapest
Weekly, online. In Hungarian.
http://www.oek.hu/oek.web?to=839&nid=41&pid=7&lang=hun

Iceland

EPI-ICE
Landlæknisembættið
Directorate Of Health, Seltjarnarnes
Monthly, online. In Icelandic and English.
http://www.landlaeknir.is

Ireland

EPI-INSIGHT
Health Protection Surveillance Centre, Dublin
Monthly, print and online. In English.
http://www.hpsc.ie/hpsc/EPI-Insight

Italy

Notiziario dell'Istituto Superiore di Sanità
Istituto Superiore di Sanità, Reparto di Malattie Infettive, Rome
Monthly, online. In Italian.
http://www.iss.it/publ/noti/index.php?lang=1&tipo=4

Bolletino Epidemiologico Nazionale (BEN)
Istituto Superiore di Sanità, Reparto di Malattie Infettive, Rome
Monthly, online. In Italian.
http://www.epicentro.iss.it/ben

Latvia

Epidemiologijas Bileteni
Sabiedribas veselibas agentura
Public Health Agency, Riga
Online. In Latvian.
http://www.sva.lv/epidemiologija/bileteni

Lithuania

Epidemiologijos žinios
Užkreciamuju ligu profilaktikos ir kontroles centras
Center for Communicable Disease Prevention and Control, Vilnius
Online. In Lithuanian.
http://www.ulac.lt/index.php?pl=26

Netherlands

Infectieziekten Bulletin
Rijksinstituut voor Volksgezondheid en Milieu
National Institute of Public Health and the Environment, Bilthoven
Monthly, print and online. In Dutch.
http://www.rivm.nl/infectieziektenbulletin

Norway

MSIS-rapport
Folkehelseinstituttet, Oslo
Weekly, print and online. In Norwegian.
http://www.folkehelsa.no/nyhetsbrev/msis

Poland

Meldunki o zachorowaniach na choroby zakazne i zatruciach w Polsce
Panstwowy Zaklad Higieny,
National Institute of Hygiene, Warsaw
Fortnightly, online. In Polish and English.
http://www.pzh.gov.pl

Portugal

Saúde em Números
Ministério da Saúde,
Direcção-Geral da Saúde, Lisbon
Sporadic, print only. In Portuguese.
http://www.dgs.pt

Slovenia

CNB Novice
Inštitut za varovanje zdravja, Center za nalezljive bolezni, Institute of Public Health, Center for Infectious Diseases, Ljubljana
Monthly, online. In Slovene.
http://www.ivz.si

Romania

Info Epidemiologia
Centrul pentru Prevenirea si Controlul Bolilor Transmisibile, National Centre of Communicable Diseases Prevention and Control, Institute of Public Health, Bucharest
Sporadic, print only. In Romanian.
Sporadic, print only. In Romanian.
http://www.insp.gov.ro/cnscbt/index.php?option=com_docman&Itemid=12

Spain

Boletín Epidemiológico Semanal
Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Madrid
Fortnightly, print and online. In Spanish.
http://revista.isciii.es

Sweden

Smittskyddsinstitutets nyhetsbrev
Smittskyddsinstitutet, Stockholm
Weekly, online. In Swedish.
http://www.smittskyddsinstitutet.se

United Kingdom

England and Wales
Health Protection Report
Health Protection Agency, London
Weekly, online only. In English.
http://www.hpa.org.uk/hpr

Northern Ireland
Communicable Diseases Monthly Report
Communicable Disease Surveillance Centre, Northern Ireland, Belfast
Monthly, print and online. In English.
http://www.cdscni.org.uk/publications

Scotland
Health Protection Scotland Weekly Report
Health Protection Scotland, Glasgow
Weekly, print and online. In English.
http://www.hps.scot.nhs.uk/ewr/

Other journals

EpiNorth journal
Norwegian Institute of Public Health, Folkehelseinstituttet, Oslo, Norway
Published four times a year in English and Russian.
http://www.epinorth.org

European Union
"Europa" is the official portal of the European Union. It provides up-to-date coverage of main events and information on activities and institutions of the European Union.
http://europa.eu

European Commission - Public Health
The website of European Commission Directorate General for Health and Consumer Protection (DG SANCO).
http://ec.europa.eu/health/

Health-EU Portal
The Health-EU Portal (the official public health portal of the European Union) includes a wide range of information and data on health-related issues and activities at both European and international level.
http://ec.europa.eu/health-eu/

Articles published in Eurosurveillance are indexed in PubMed/Medline.

The Eurosurveillance print edition is a selection of short and long articles previously published on the Eurosurveillance website. The full listing of all Eurosurveillance articles can be found in the Archives section of the website.

TQ-AD-13-001-EN-C

Eurosurveillance

Visit our website at **www.eurosurveillance.org**

The Eurosurveillance print edition is a compilation of short and long articles that have previously been published on our website.

All the articles in this issue are available online: you can print each page separately or download the whole quarterly in pdf format.

The website archives all articles since 1995, and offers a search facility.

To receive Eurosurveillance's free **electronic releases** and e-alerts by e-mail, please subscribe on our website.

Papers published in the former monthly release are indexed for MEDLINE since January 2001, and papers published in the weekly release from January 2005 (with the exception of short, non-scientific notices) are also indexed for MEDLINE.

The Index Medicus abbreviation for Eurosurveillance is Euro Surveill.

Follow us on Twitter!
https://twitter.com/#!/eurosurveillanc

Contributions to Eurosurveillance are welcomed. Full instructions to authors are available at our website www.eurosurveillance.org

ISSN 1025 496X
4,800 copies
Graphic design © ECDC, Stockholm

Publications Office