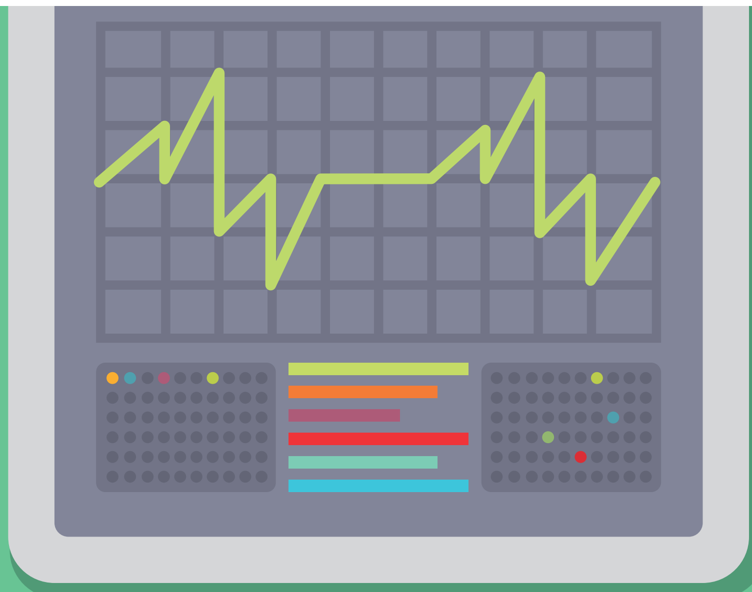
Impact factor
7.1

Eurosurveillance

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



Special edition:

Advanced diagnostics to inform public health policy

March 2019

Guest editor: Jacob Moran-Gilad

Featuring

- How do advanced diagnostics support public health policy development?
- Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157:H7, England, 2013 to 2017
- and more...

ecdc EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

www.eurosurveillance.org

Editorial team

Based at the European Centre for
Disease Prevention and Control (ECDC),
169 73 Stockholm, Sweden

Telephone number

+46 (0)8 58 60 11 38

E-mail

eurosurveillance@ecdc.europa.eu

Editor-in-chief

Dr Ines Steffens

Senior editor

Kathrin Hagmaier

Scientific editors

Janelle Sandberg

Karen Wilson

Assistant editors

Alina Buzdugan

Associate editors

Tommi Asikainen, Brussels, Belgium

Magnus Boman, Stockholm, Sweden

Mike Catchpole, Stockholm, Sweden

Natasha Crowcroft, Toronto, Canada

Christian Drosten, Berlin, Germany

Karl Ekdahl, Stockholm, Sweden

Johan Giesecke, Stockholm, Sweden

David Heymann, London, United Kingdom

Irena Klavs, Ljubljana, Slovenia

Karl Kristinsson, Reykjavik, Iceland

Daniel Lévy-Bruhl, Paris, France

Jacob Moran-Gilad, Beer-Sheva, Israel

Chantal Reusken, Bilthoven, the Netherlands

Panayotis T. Tassios, Athens, Greece

Hélène Therre, Paris, France

Henriette de Valk, Paris, France

Sylvie van der Werf, Paris, France

Design / Layout

Fabrice Donguy / Dragos Platon

Online submission system

http://www.editorialmanager.com/eurosurveillance/

www.eurosurveillance.org

© Eurosurveillance, 2019

Editorial advisors

Albania: Alban Ylli, Tirana

Austria: Maria Paulke-Korinek, Vienna

Belgium: Koen de Schrijver, Antwerp; Tinne Lernout, Brussels

Bosnia and Herzegovina: Nina Rodić Vukmir, Banja Luka

Bulgaria: Iva Christova, Sofia

Croatia: Sanja Music Milanovic, Zagreb

Cyprus: Maria Koliou, Nicosia

Czech Republic: Jan Kynčl, Prague

Denmark: Peter Henrik Andersen, Copenhagen

Estonia: Kuulo Kutsar, Tallinn

Finland: Outi Lyytikäinen, Helsinki

France: Judith Benrekassa, Paris

Germany: Jamela Seedat, Berlin

Greece: Rengina Vorou, Athens

Hungary: Ágnes Hajdu, Budapest

Iceland: Gudrun Sigmundsdottir, Reykjavík

Ireland: Joan O Donnell, Dublin

Italy: Paola De Castro, Rome

Latvia: Dzintars Mozgis, Riga

Lithuania: Saulius Čaplinskas, Vilnius

Luxembourg: Thérèse Staub, Luxembourg

The former Yugoslav Republic of Macedonia: Aziz Pollozhani, Skopje

Malta: Tanya Melillo Fenech, Msida

Montenegro: Senad Begić, Podgorica

Netherlands: Barbara Schimmer, Bilthoven

Norway: Emily MacDonald, Oslo

Poland: Malgorzata Sadkowska-Todys, Warsaw

Portugal: Paulo Jorge Nogueira, Lisbon

Romania: Daniela Pitigoi, Bucharest

Serbia: Mijomir Pelemis, Belgrade

Slovakia: Lukáš Murajda, Bratislava

Slovenia: Maja Sočan, Ljubljana

Spain: Josefa Masa Calle, Madrid

Sweden: Anders Wallensten, Stockholm

Turkey: Fehminaz Temel, Ankara

United Kingdom: Nick Phin, London

World Health Organization Regional Office for Europe:
Masoud Dara, Copenhagen

Contents

Special edition: Advanced diagnostics to inform public health policy

Editorial

How do advanced diagnostics support public health policy development? 2
Moran-Gilad J

Surveillance and outbreak reports

Whole genome sequencing–based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations 6
Abascal et al.

Cocirculation of Hajj and non-Hajj strains among serogroup W meningococci in Italy, 2000 to 2016 20
Fazio et al.

Research Articles

Culture-free genotyping of *Neisseria gonorrhoeae* revealed distinct strains at different anatomical sites in a quarter of patients, the Netherlands, 2012 to 2016 29
Van der Veer et al.

Improvement of Legionnaires' disease diagnosis using real-time PCR assay: a retrospective analysis, Italy, 2010 to 2015 38
Ricci M L et al.

Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples 43
Kafetzopoulou L E et al.

Prospective genomic surveillance of methicillin-resistant *Staphylococcus aureus* (MRSA) associated with bloodstream infection, England, 1 October 2012 to 30 September 2013 56
Toleman M et al.

Whole genome sequencing of *Salmonella* Chester reveals geographically distinct clusters, Norway, 2000 to 2016 68
Siira L et al.

Air-conditioner cooling towers as complex reservoirs and continuous source of Legionella pneumophila infection evidenced by a genomic analysis study in 2017, Switzerland 79
Wüthrich D et al.

Review

Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018 86
Rodriguez-Sanchez B et al.

Perspective

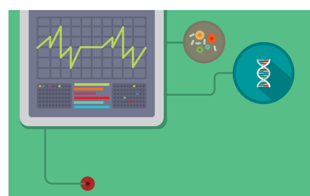
Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup O157:H7, England, 2013 to 2017 98
Jenkins C et al.

Guest editor: Jacob Moran-Gilad, Ministry of Health, Jerusalem and Ben-Gurion University of the Negev, Beer-Sheva, Israel

Professor Jacob Moran-Gilad is a board-certified consultant clinical microbiologist and in public health / medical epidemiology. He is the principal investigator of the MAGICAL group (Microbiology, Advanced Genomics and Infection Control Application Laboratory) at the Faculty of Health Sciences, Ben-Gurion University (BGU) and the Co-Chair of the BGU Research Hub for Infectious Diseases and One Health. His current research involves the development and validation of laboratory methods for rapid diagnosis of infectious diseases, the study of the molecular epidemiology of various pathogens of public health interest and the integration and application of new methods in health protection, using advanced genomics and metagenomics. He is also involved in research on public health microbiology policy and planning, and in exploring contemporary approaches for linking up medical epidemiology and cutting-edge microbiology to inform public health strategy.

As a senior consultant to Israeli Ministry of Health, he is focusing on national strategy and policy-making in public health microbiology. Under this capacity he is the Chair of several national infectious disease programmes, involved in emergency preparedness and response and chairs the National Advisory Committee for Microbiology. Professor Moran-Gilad is also the lead Clinical Microbiologist at Soroka University Medical Center, a 1,200-bed tertiary-care busy regional medical centre in Southern Israel and oversees the microbiological diagnostic activities in that institution.

At the international level, Professor Moran-Gilad is the Chairperson of the ESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD), member of the board of the ESCMID Study Group for *Legionella* Infections (ESGLI) responsible for microbiology and genomics and Chair of the International Working Group for whole genome sequencing of Legionella. Recently, he was appointed as the Deputy Programme Director and the designate Programme Director of ECCMID, the world's leading conference on microbiology and infection. Jacob is as an Associate editor in the *Eurosurveillance* editorial board.



© Istockphoto

Editorial

How do advanced diagnostics support public health policy development?

Jacob Moran-Gilad[1,2]
1. Dept. of Health Systems Management, School of Public Health, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel
2. ESCMID Study Group for Genomic and Molecular Diagnostics, Basel, Switzerland

Correspondence: Jacob Moran-Gilad (giladko@post.bgu.ac.il)

Citation style for this article:
Moran-Gilad Jacob. How do advanced diagnostics support public health policy development?. Euro Surveill. 2019;24(4):pii=1900068. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1900068

Article submitted on 23 Jan 2019 / accepted on 24 Jan 2019 / published on 24 Jan 2019

Microbiologists working in clinical/diagnostic microbiology or public health microbiology (mainly food, water and environmental), have experienced a major revolution of their profession over recent years. Technological advancements involving the development and implementation of new analytical platforms have allowed for faster, more accurate and more complex diagnostics [1]. Some of these technologies are novel and emerge as ‘disruptive technologies’, while others improve and enhance existing diagnostic approaches. In this context, how do we define ‘advanced diagnostics’?

Advanced diagnostics can be divided into several groups, according to their methodological approach as well as their practical applications. One such division differentiates between culture-dependent (culture-based) and culture-independent microbiology (Table). With culture-based diagnostics, applicable mainly to bacterial and fungal pathogens, one or more culture phases are involved in order to yield growth of the suspected microorganism from a clinical or non-clinical sample. Subsequently, growing isolates are characterised with respect to taxonomy, antimicrobial drug susceptibility and other traits (such as virulence and molecular subtypes) by a range of approaches. These mainly include—but are not necessarily restricted to—characterisation by conventional (phenotypic) techniques, molecular assays targeting specific genes, proteomics (primarily taxonomical identification using matrix-assisted laser desorption-ionisation time-of-flight mass spectrometry (MALDI-TOF-MS)) or single-cell whole genome sequencing (WGS), followed by bioinformatics analyses to call the taxonomy and phylogenomic subtype and infer phenotypic resistance and virulence, by mapping the resistome and virulome. WGS, powered by next-generation sequencing (NGS), is undoubtedly the most impactful application, downstream to culture isolation, and has the potential to serve as a one-stop-shop for pathogen characterisation, while allowing for unprecedented accuracy and resolution [2].

On the other hand, culture-independent microbiology involves the application of diagnostic techniques directly on clinical or non-clinical samples, while obviating the need to recover an organism by culture. This approach has long been used in the field of virology, where virus isolation is rarely performed for routine diagnostic purposes whereas it was not common practice for other pathogens. However, culture-independent detection methods are also applicable to bacterial, fungal and parasitic diseases. With culture-independent microbiology, several diagnostic strategies are now commonly used also for the latter group of pathogens, including the application of PCR assays targeting specific genes that relate to presence of a pathogen and/or an important inferred phenotype, such as antimicrobial resistance to a key agent. More recently, a massive increase in the availability of in-house and commercial multiplex PCR assays is evident, covering a wide range of diagnostic targets in a single run. These assays are increasingly designed for syndromic diagnosis, covering the most common pathogens causing infection in well-defined infectious disease syndromes such as respiratory, gastrointestinal or genitourinary syndromes, as well as syndromes caused by central nervous system infections and even bloodstream infections [3]. Rapid diagnostic tests (RDTs) that are derivatives of syndromic multiplex assays have been designed to generate rapid results in a fairly robust manner and they could be used outside the medical laboratory, closer to the patient or in the field, even by non-laboratorians [4]. These point of care (POC) or point of impact (POI) molecular tests are highly promising also with respect to their impact on public health. Lastly, applying NGS technology directly on samples, an approach also known as metagenomics, has been used for many years now in ecology and environmental

Table

Advanced diagnostics by technology and approaches, 2019

Approach	Technology							
	Conventional / standard microbiology	Molecular microbiology		Proteomics	Molecular	Genomics / metagenomics		
		PCR	Multiplex PCR	MALDI-TOF-MS	standard typing methods	WGS	Microbiomics	Whole genome metagenomics
Culture-based	Organism ID/AST	Detection/Sanger sequencing of specific gene for characterisation of grown organism (e.g. resistance or virulence determinant)	Detection of specific genes for characterisation of grown organism (e.g. resistance or virulence determinant),	Identification of grown organism; more recently, potential for detection of resistance or typing	PFGE, SLST, MLST, MLVA	ID/AST, mapping of resistome and virulome, typing by SNPs or cgMLST	NA	NA
Culture-independent	NA	Detection of specific genes, for organism presence (or characteristic such as presence of specific gene)	Syndromic testing for a range of potential pathogens per sample type	Application of MALDI-TOF-MS directly on samples still experimental	NA	NA	Microbial population analysis	Microbial population analysis, functional characterisation, extraction of whole genome assemblies, phenotype prediction

AST: antimicrobial susceptibility testing; cgMLST: core genome multilocus sequence typing; ID: identification; MALDI-TOF MS: matrix-assisted laser desorption ionization-time of flight mass spectrometry: MLST: multilocus sequence typing; MLVA: multilocus variable number tandem repeat analysis; NA: not applicable; PFGE: pulsed-field gel electrophoresis; SLST: singlelocus sequence typing; SNP: single nucleotide polymorphism; WGS: whole genome sequencing.

sciences. It has the potential, when applied on clinical materials, to accurately map the microbial population in a body site (i.e. the microbiome) by amplification of a target gene such as the 16S rRNA gene, or to generate information regarding the entire taxonomical composition of a sample, while allowing deeper analysis of microbial characteristics and functions (shotgun or whole genome metagenomics) [5]. The latter is especially appealing because of its potential for not only analysing the microbiota, but also allowing whole genome assemblies' extraction from the metagenome, enabling therapeutic inferences and, in the future, complementary analysis of the host human genome or transcriptome for tailoring treatment and establishing prognosis.

In this special issue of *Eurosurveillance*, 10 articles describe the development and application of such advanced diagnostics, with respect to communicable diseases of public health concern. Through this suite of articles, it is evident that the diagnostic revolution in the field of microbiology is already creating a major impact on public health response and policy making related to infectious diseases.

Two papers focus on harnessing WGS for performing national surveillance of pathogens of public health importance. The first, by Toleman et al., demonstrates the added value of genomic surveillance of meticillin-resistant *Staphylococcus aureus* (MRSA) in the United Kingdom (UK) [6]. This one-year study of all available isolates implicated in bloodstream infections demonstrated the dynamics of MRSA diversity in the UK, identified high-risk clones and contextualised several reported outbreaks. The second paper, by Jenkins et al., shares the UK experience of standardising genomic surveillance of Shiga-toxin producing *Escherichia coli* (STEC) as a foodborne pathogen [7]. This effort proved successful with respect to resolving case clusters with obscure epidemiological data and provided insight into the evolution of pathogenic strain and geographical spread.

Four papers focus on employing WGS for cluster/outbreak investigation in different settings. Fazio et al. studied the increase in serogroup W *Neisseria meningitidis* in Italy over nearly two decades, showing an unusual cocirculation of two meningococcal lineages originating from South America and the Hajj pilgrimage [8]. Similarly, Siira et al. investigated an increase in *Salmonella* Chester infections in Norway also over nearly two decades. WGS dissected this cluster of cases into several distinct geographical origins and unravelled the occurrence of an outbreak originating in another European country [9]. Abascal et al. used WGS to target cross-border surveillance of tuberculosis in Spain. Their data confirm the limitations of the mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) approach, in that MIRU-VNTR failed to discriminate importations and recent transmissions [10]. Finally, Wüthrich et al. studied an exceedance of legionellosis cases in the city of Basel, Switzerland. Genomic analysis revealed several interesting features, including the contamination of cooling towers by multiple strains, the involvement of highly conserved strains in causing disease over a long time period and the interrelations between cooling towers, which could form a complex microbial network in the same area [11].

Rodriguez-Sánchez et al. reviewed the utility of MALDI-TOF-MS for public health purposes, beyond the main application of proteomics. Such applications include direct application of MALDI-TOF MS on positive blood cultures to improve time to detection of pathogens causing bacteraemia (especially Gram-negative rods), using MALDI-TOF-MS for identification of molecular mechanisms of resistance such as carbapenemases and using MALDI-TOF MS for phylogenetic typing for strains tracking and outbreak detection [12].

Three papers demonstrate the strength of culture-independent microbiology. Ricci et al. performed an evaluation of a commercial and an in-house qPCR assay for the detection of *Legionella pneumophila* in respiratory samples [13]. Their results show that qPCR outperformed the urinary antigen test and culture. While these findings are not unexpected, mindful of the known limitation of these two methods, the increase in sensitivity by molecular diagnosis has public health implications, as more Legionnaires' disease cases and clusters will be detected and investigated. In another paper, van der Veer et al. report on a culture-independent method they developed for typing *Neisseria gonorrhoeae* [14]. This approach is advantageous, as typing of this fastidious organism requires its isolation in culture, which may be challenging. The method developed and implemented by the authors improved the typeability by ca 50%. Interestingly, this approach has also shown that multiple subtypes may coinfect individuals, which is an important epidemiological finding that would have otherwise been missed, should culture be performed as per existing guidelines from a single anatomical site. Lastly, Kafetzopoulou et al. have used metagenomics to recover the near-full sequences of arboviruses from clinical samples that tested positive for chikungunya or dengue viruses using real-time reverse transcription-PCR [15]. The authors have successfully used two different sequencing technologies. While the samples sequenced were serum/plasma, which are normally sterile, making the bioinformatics analysis for genome recovery less challenging, these findings are encouraging with respect to the feasibility of future metagenomics approaches for arboviral diseases.

Despite the promising results, several challenges remain and need to be addressed by the public health, microbiological and infectious disease communities. Reliance on culture-based methods prolongs the turnaround time for diagnosis and, despite WGS being increasingly streamlined, producing clinically actionable information in real-time via WGS is still challenging. Moreover, predicting phenotypes based on genomics (e.g. prediction of minimum inhibitory concentration to antimicrobials) is still not readily achievable [16]. MALDI-TOF MS has become very popular and many frontline laboratories are using it routinely. Still, more advanced applications of MALDI-TOF MS, such as assessment of antimicrobial resistance or typing, require more development and validation [16]. With culture-independent approaches, multiplex testing may detect non-culturable, non-viable organisms whose significance is unknown, as is the frequent detection of co-infections that are difficult to translate into management decisions while validation is ongoing. Increased reliance on multiplex PCRs also suggests the reduced availability of cultured organisms, which has consequences with respect to strain referral and reference microbiology as a central element of microbiological surveillance at national and international levels. With metagenomics there are still many hindrances, including costs, disparities in capabilities and capacities for performing deep sequencing, optimisation of sample preparation and, most importantly, the bioinformatics analysis, which is incredibly complex, especially when genotype to phenotype correlations are sought.

As proteomics, genomics and metagenomics are increasingly being implemented in microbiology laboratories there are many aspects that need further consideration. These encompass quality control, including the use of certified reference materials and internal and external quality assurance [1,17,18]. Furthermore, there is a need for validation of bioinformatics pipelines that will allow a standardised analysis [19] and meet accreditation requirements, for ensured reverse compatibility between methods [18], for data safety and security, for data sharing agreements as well as deposition and metadata collection etc. The successful implementation of advanced diagnostics in the service of public health, thus depends on many factors. Appropriate national and international frameworks are needed that support timely diagnosis of infectious diseases and high pathogen resolution by using the most appropriate diagnostic methods available today or becoming available in the near future.

Acknowledgements

JMG conceptualised and acted as guest editor for the *Eurosurveillance* special issue on advanced diagnostics to inform public health policy. He is an associate editor of *Eurosurveillance*.

Conflict of interest

None declared.

References

1. Motro Y, Carriço JA, Friedrich AW, Rossen JWA, Moran-Gilad J. ESCMID postgraduate education course: regional capacity building for integration of next-generation sequencing in the clinical microlab. Microbes Infect. 2018;20(5):275-80. https://doi.org/10.1016/j.micinf.2018.02.006 PMID: 29555344
2. Motro Y, Moran-Gilad J. Next-generation sequencing applications in clinical bacteriology. Biomol Detect Quantif. 2017;14:1-6. https://doi.org/10.1016/j.bdq.2017.10.002 PMID: 29255684
3. Ramanan P, Bryson AL, Binnicker MJ, Pritt BS, Patel R. Syndromic Panel-Based Testing in Clinical Microbiology. Clin Microbiol Rev. 2017;31(1):e00024-17. https://doi.org/10.1128/CMR.00024-17 PMID: 29142077
4. van Belkum A, Bachmann TT, Lüdke G, Lisby JG, Kahlmeter G, Mohess A, et al. JPIAMR AMR-RDT Working Group on

Antimicrobial Resistance and Rapid Diagnostic Testing. Developmental roadmap for antimicrobial susceptibility testing systems. Nat Rev Microbiol. 2019;17(1):51-62. https://doi.org/10.1038/s41579-018-0098-9 PMID: 30333569

5. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35(9):833-44. https://doi.org/10.1038/nbt.3935 PMID: 28898207
6. Toleman MS, Reuter S, Jamrozy D, Wilson H, Blane B, Harrison EM, et al. Prospective genomic surveillance of methicillin-resistant Staphylococcus aureus (MRSA) associated with bloodstream infection, England, 1 October 2012 to 30 September 2013. Euro Surveill. 2019;24(4):1800215. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800215
7. Jenkins C, Dallman TJ, Grant KA. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup O157:H7, England, 2013 to 2017. Euro Surveill. 2019;24(4):1800346. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800346
8. Fazio C, Neri A, Vacca P, Ciammaruconi A, Arghittu M, Barbui AM, et al. Cocirculation of Hajj and non-Hajj strains among serogroup W meningococci in Italy, 2000 to 2016. Euro Surveill. 2019;24(4):1800183. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800183
9. Siira L, Naseer U, Alfsnes K, Hermansen NO, Lange H, Brandal LT. Whole genome sequencing of Salmonella Chester reveals geographically distinct clusters, Norway, 2000 to 2016. Euro Surveill. 2019;24(4):1800186. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800183
10. Abascal E, Pérez-Lago L, Martínez-Lirola M, Chiner-Oms Á, Herranz M, Chaoui I, et al. Whole genome sequencing–based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. Euro Surveill. 2019;24(4):1800005. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800005
11. Wüthrich D, Gautsch S, Spieler DR, Dubuis O, Gaia V, Moran-Gilad J, et al. Air-conditioner cooling towers as complex reservoirs and continuous source of Legionella pneumophila infection evidenced by a genomic analysis study in 2017, Switzerland. Euro Surveill. 2019;24(4):1800192. .
12. Rodriguez-Sanchez B, Cercenado E, Coste AT, Greub G. Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018. Euro Surveill. 2019;24(4):1800193. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800193
13. Ricci ML, Grottola A, Fregni Serpini G, Bella A, Rota MC, Frascaro F, et al. Improvement of Legionnaires' disease diagnosis using real-time PCR assay: a retrospective analysis, Italy, 2010 to 2015. Euro Surveill. 2018;23(50):1800032. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800032 PMID: 30563592
14. van der Veer BMJW, Wolffs PFG, Hoebe CJPA, Dukers-Muijrers NHTM, van Alphen LB. Culture-free genotyping of Neisseria gonorrhoeae revealed distinct strains at different anatomical sites in a quarter of patients, the Netherlands, 2012 to 2016. Euro Surveill. 2018;23(50):1800253. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800253 PMID: 30563596
15. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, Crook A, Carter D, Osborne J, et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Euro Surveill. 2018;23(50):1800228. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800228 PMID: 30563591
16. Greub G, Moran-Gilad J, Rossen J, Egli AESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD). ESCMID postgraduate education course: applications of MALDI-TOF mass spectrometry in clinical microbiology. Microbes Infect. 2017;19(9-10):433-42. https://doi.org/10.1016/j.micinf.2017.06.004 PMID: 28669790
17. Moran-Gilad J, Sintchenko V, Pedersen SK, Wolfgang WJ, Pettengill J, Strain E, et al. Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC Infect Dis. 2015;15(1):174. https://doi.org/10.1186/s12879-015-0902-3 PMID: 25887164
18. Rossen JWA, Friedrich AW, Moran-Gilad JESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD). Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. Clin Microbiol Infect. 2018;24(4):355-60. https://doi.org/10.1016/j.cmi.2017.11.001 PMID: 29117578
19. Carriço JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. Clin Microbiol Infect. 2018;24(4):342-9. https://doi.org/10.1016/j.cmi.2017.12.015 PMID: 29309933

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Surveillance

Whole genome sequencing–based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations

Estefanía Abascal[1,2,3], Laura Pérez-Lago[1,2,3], Miguel Martínez-Lirola[4], Álvaro Chiner-Oms[5], Marta Herranz[1,2,6], Imane Chaoui[7], Iñaki Comas[8,9], My Driss El Messaoudi[10], José Antonio Garrido Cárdenas[11], Sheila Santantón[1,2], Emilio Bouza[1,2,6,12], Darío García-de-Viedma[1,2,6]

1. Servicio Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón, Madrid, Spain
2. Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
3. These authors have contributed equally
4. Complejo Hospitalario Torrecárdenas, Almería, Spain
5. Unidad Mixta Genómica y Salud, Centro Superior de Investigación en Salud Pública (FISABIO)-Universitat de València, Valencia, Spain
6. CIBER Enfermedades respiratorias (CIBERES), Spain
7. Unité de Biologie et Recherches Médicales, Division des Sciences du Vivant, Centre National de l'Energie, des Sciences et des Techniques Nucléaires (CNESTEN), Rabat, Morocco
8. Instituto de Biomedicina de Valencia (IBV) Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain
9. CIBER Epidemiología y Salud Pública (CIBERESP), Spain
10. Institut Pasteur of Morocco, Casablanca, Morocco
11. Servicio de Secuenciación, Universidad de Almería, Almería, Spain
12. Departamento de Medicina, Facultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain

Correspondence: Darío García de Viedma (dgviedma2@gmail.com)

Citation style for this article:
Abascal Estefanía, Pérez-Lago Laura, Martínez-Lirola Miguel, Chiner-Oms Álvaro, Herranz Marta, Chaoui Imane, Comas Iñaki, El Messaoudi My Driss, Cárdenas José Antonio Garrido, Santantón Sheila, Bouza Emilio, García-de-Viedma Darío. Whole genome sequencing–based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. Euro Surveill. 2019;24(4):pii=1800005. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800005

Article submitted on 26 Dec 2017 / accepted on 01 Jun 2018 / published on 24 Jan 2019

Background: The analysis of transmission of tuberculosis (TB) is challenging in areas with a large migrant population. Standard genotyping may fail to differentiate transmission within the host country from new importations, which is key from an epidemiological perspective. **Aim:** To propose a new strategy to simplify and optimise cross-border surveillance of tuberculosis and to distinguish between recent transmission in the host country and new importations. **Methods:** We selected 10 clusters, defined by 24-locus mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR), from a population in Spain rich in migrants from eastern Europe, north Africa and west Africa and reanalysed 66 isolates by whole-genome sequencing (WGS). A multiplex-allele-specific PCR was designed to target strain-specific marker single nucleotide polymorphisms (SNPs), identified from WGS data, to optimise the surveillance of the most complex cluster. **Results:** In five of 10 clusters not all isolates showed the short genetic distances expected for recent transmission and revealed a higher number of SNPs, thus suggesting independent importations of prevalent strains in the country of origin. In the most complex cluster, rich in Moroccan cases, a multiplex allele-specific oligonucleotide-PCR (ASO-PCR) targeting the marker SNPs for the transmission subcluster enabled us to prospectively identify new secondary cases. The ASO-PCR-based strategy was transferred and applied in Morocco, demonstrating that the strain was prevalent in the country. **Conclusion:** We provide a new model for optimising the analysis of cross-border surveillance of TB transmission in the scenario of global migration.

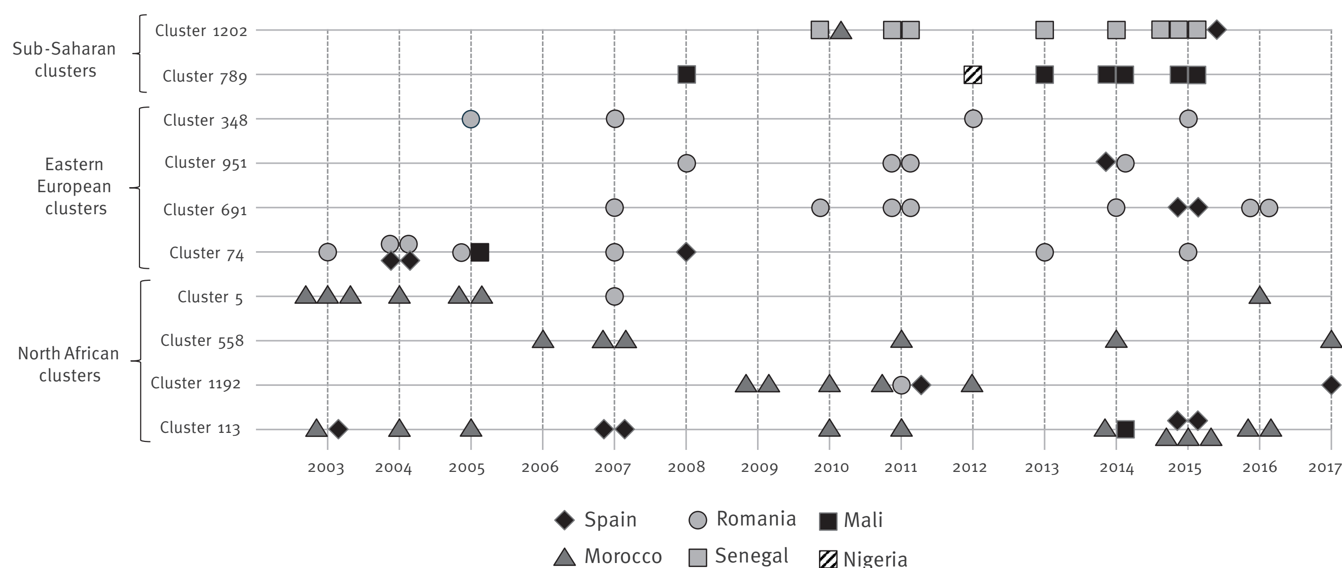
Background

International migration has modified the epidemiology of tuberculosis (TB) in most high-income countries and today, migrants account for up to 40–60% of cases in large cities [1-4]. Some cases are reactivations of infections acquired in the country of origin, with the remainder resulting from recent transmission after arrival in the host country.

Molecular epidemiology provides more accurate data on the transmission dynamics of TB in settings with a complex composition of cases due to migration [5-7]. Several studies have shown variable composition in the nationalities comprising transmission clusters. This variety ranges from settings with marked transmission permeability leading to multinational clusters,

Figure 1

Chart summarising the general data of the clusters analysed, rich in cases from sub-Saharan Africa, eastern Europe and north Africa, 2003–2017 (n = 10 clusters)



Clusters are grouped according to their geographic origin (sub-Saharan, eastern European and north African). Each horizontal line corresponds to a cluster and each symbol corresponds to a patient. The patients involved in each cluster are distributed along the timeline (years at the bottom of the chart) with different symbols according to the nationalities shown in the legend.

to other socio-epidemiological contexts where a more homogeneous composition of nationalities is found, with clusters only involving single nationalities [6,8]. Autochthonous clusters and those comprising several nationalities more likely reflect recent transmission events. However, clusters rich in cases from one country of origin are especially difficult to interpret. This is because they can be the result of either of two circumstances: (i) a strain is imported from the country of origin and subsequently transmitted to migrants of the same nationality in the host country; or (ii) genetically closely related strains, which are prevalent in the country of origin, are independently imported by individuals who were exposed in the country of origin but are not epidemiologically related in the host country. Thus, differentiation between these alternatives, i.e. recent transmission in the host country vs importation, is challenging, yet highly relevant in epidemiological terms.

Application of whole-genome sequencing (WGS) for analysis of transmission of TB has given birth to the field of genomic epidemiology, which has markedly increased specificity in the definition of transmission clusters [9-12]. Determination of the number of single nucleotide polymorphisms (SNPs) [12] between the sequences of different isolates allows to split clusters that had been previously defined by standard molecular tools into smaller subclusters that are much more consistent with the geographic distribution of the cases and with the epidemiological links between them [11].

Our aim was to apply WGS in a more in-depth analysis of migrant TB cases involved in clusters in Spain that had been defined by standard genotyping. We attempted to determine whether the clusters corresponded to recent transmission in the host country (because *Mycobacterium tuberculosis* (MTB) isolates show no or a very short genetic distance) or to undetected independent importations of strains that are prevalent in the country of origin and have acquired higher SNP-based diversity as a result of prolonged periods of circulation. In addition, we took advantage of the SNPs identified for either the recently transmitted or imported isolates, to tailor simple PCR tools to simplify and optimise the precise assignation of recent transmission or importation in the new clusters arising. Further, we used these same tools in a new extended and cross-border analysis, for an in-depth surveillance of the MTB strains analysed in unrelated Spanish populations, as well as in the country of origin.

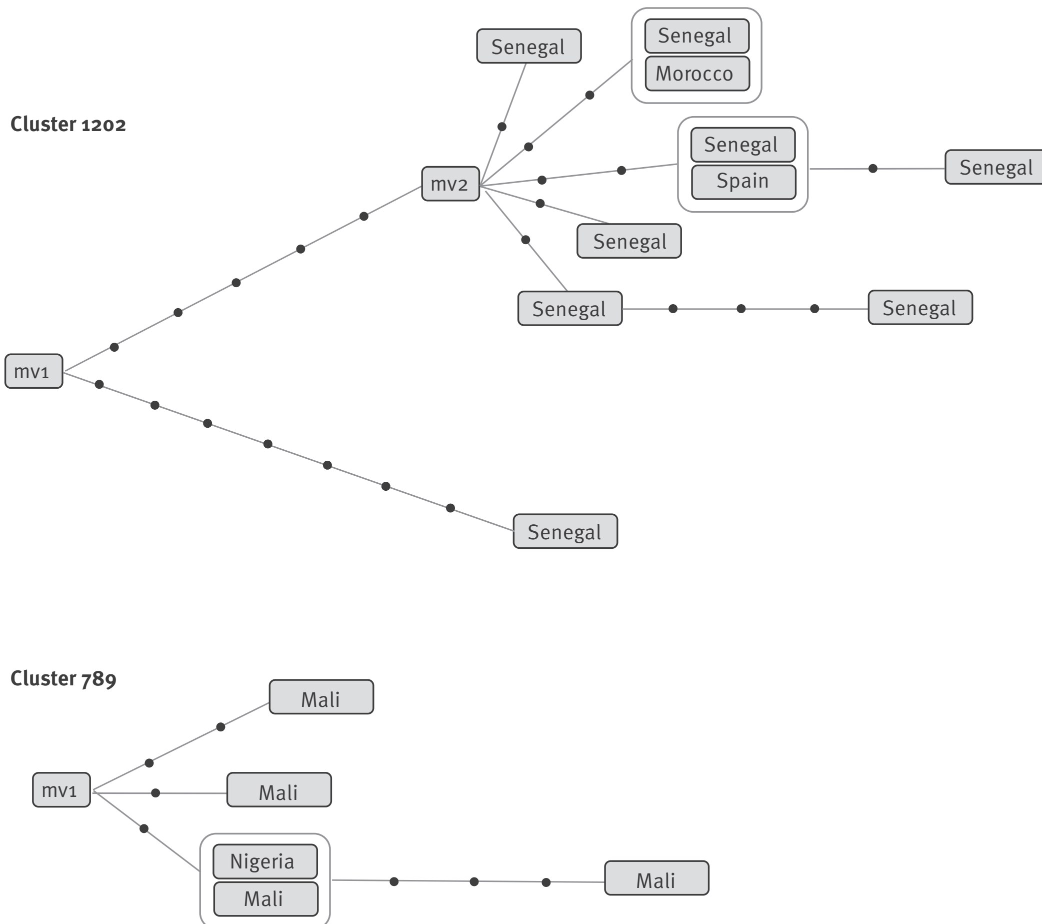
Methods

Clusters and strains selected

We retrospectively selected all clusters from the ongoing molecular epidemiology universal genotyping programme in Almería, south-east Spain [7,13] fulfilling the following selection criteria: The clusters analysed were 24 locus mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR)–defined clusters [14] including four or more cases, covering at least 5 years and rich (>60% of the clustered cases) in migrants from a single country from one of

Figure 2

Networks of relationships obtained from the whole genome sequencing analysis for clusters rich in cases from sub-Saharan Africa



mv: median vectors; SNP: single nucleotide polymorphisms.

Each black dot corresponds to a SNP. Each box corresponds to a patient and the nationality is indicated within the box. When two or more cases share identical sequences (zero SNPs between them) they are surrounded by a line.

three geographic areas (eastern Europe, north Africa and sub-Saharan Africa). The lineage of the strains involved in the selected clusters was assigned based on the determination of lineage-specific SNP markers [15] by multiplex allele-specific oligonucleotide-PCR (ASO-PCR) [16].

Convenience samples from Valencia (all isolates with available WGS data in IBV, for the period 2004-2017) and Madrid (all isolates with genotypic data available in Hospital Gregorio Marañón, Spain) for the period 2004-10, were also included in the study. A retrospective convenience sample of part of the isolates from northern Morocco (Tangier, Tetouan and Larache) obtained during the same period also were included; no previous genotypic information was available for these isolates. Finally, a pool of 20 randomly selected TB migrant cases from Morocco (among all those diagnosed in Almería) that were infected with strains other than those analysed in this study were selected as controls.

Genomic analysis

DNA purification

DNA for WGS of the MIRU-VNTR-defined clusters from Almería was purified from subcultures on Mycobacteria Growth Indicator Tube (MGIT) (using Qiagen kit; QIAamp DNA Mini Kit, Qiagen, Courtaboeuf, France) or Lowenstein Jensen medium (CTAB (cetyl trimethylammonium bromide)-based standard purification).

WGS of the strains from the collection in Morocco was performed by purifying (Qiagen kit) the DNA from the remnants of bacterial lysates that had been stored.

WGS of the strains from the collection in Madrid was performed by purifying DNA (Qiagen kit) from freshly inactivated suspensions from the stored frozen isolates.

Whole genome sequencing and single nucleotide polymorphism analysis

WGS was performed as detailed elsewhere [17]. Briefly, DNA libraries were generated following the Nextera XT Illumina protocol (Nextera XT Library Prep kit (FC-131–1024), Illumina, San Diego, United States (US)). Library quality and size distribution were checked on a 2200 TapeStation Bioanalyzer (Agilent Technologies, Santa Clara, US). Libraries were run in a Miseq device (Illumina), which generated 35–151–bp paired-end reads and an average per base coverage of 70 x. Sequences were deposited in www.ebi.ac.uk (PRJEB23664 and PRJEB25814).

We mapped the reads for each strain using the Burrows-Wheeler Aligner and the ancestral MTB genome, which was identical to H37Rv in terms of structure, but which included the maximum likelihood–inferred ancestral nt positions from a virtual ancestor [18]. SNP calls were made with SAMtools and VarScan (coverage of at least 20 x, mean SNP mapping quality of 20). From all the variants detected, we kept only the homozygous calls (those present in at least 90% of the reads in a specific position). Moreover, to filter out potential false positive SNPs due to mapping errors we omitted the variants detected in repetitive regions, phages and PE/PPE regions. Also, SNPs close to indels and those present in areas with an anomalous accumulation of variants (three or more SNPs in 10 bp) were omitted. Alignments and SNP variants (called with a › 20 x coverage in at least one of the isolates in a cluster) were visualised and checked for the remaining isolates in the Integrative Genomics Viewer IGV (version 2.3.59) programme. Multiple comparisons between the SNPs from different isolates were made using an in-house script written in R [19]. We used the reference values (in the number of SNPs) of Walker et al. [12] to determine whether the isolates in a MIRU-VNTR cluster were related. In three isolates we detected an unexpectedly high number of SNPs (› 200) with respect to the other members in the cluster; they were considered to be clustered as the result of homoplasy in the MIRU-VNTR pattern and therefore were eliminated from the study.

The median-joining networks were constructed from the SNP matrix generated for each case using the programme NETWORK 5.0.0.1. Median vectors (mv) were defined when the distribution of SNPs of the isolates analysed indicated the existence of a node that was not represented by the sampled isolates sequenced for each cluster. These median vectors therefore corresponded to non-sampled isolates in the cluster. The chronology of acquisition of SNPs is represented from left to right in the networks.

Cluster-specific single nucleotide polymorphisms and design of ASO-PCRs

To identify SNPs which were specific for cluster 113, we created a database of variants using sequences from isolates which were representative of the global MTB complex (MTBC) diversity. We downloaded all the accessible raw data from different publications [20-22]. All the fastq files published in these studies were downloaded and aligned against the ancestral MTB genome using the BWA tool. We kept the alignments that had a mean coverage higher than 20. Using this criterion, we kept 7,977 samples representative from the seven lineages. We extracted all the variants present in these samples as described above. The 7,977 samples were filtered to remove transmission clusters so we kept one representative strain of each transmission cluster detected. Once the transmission clusters were filtered, we kept 4,762 sequences. The 207,188 variants present in these samples were used to construct a reference database to evaluate the specificity of the SNPs selected for the ASO-PCRs to be applied in cluster 113.

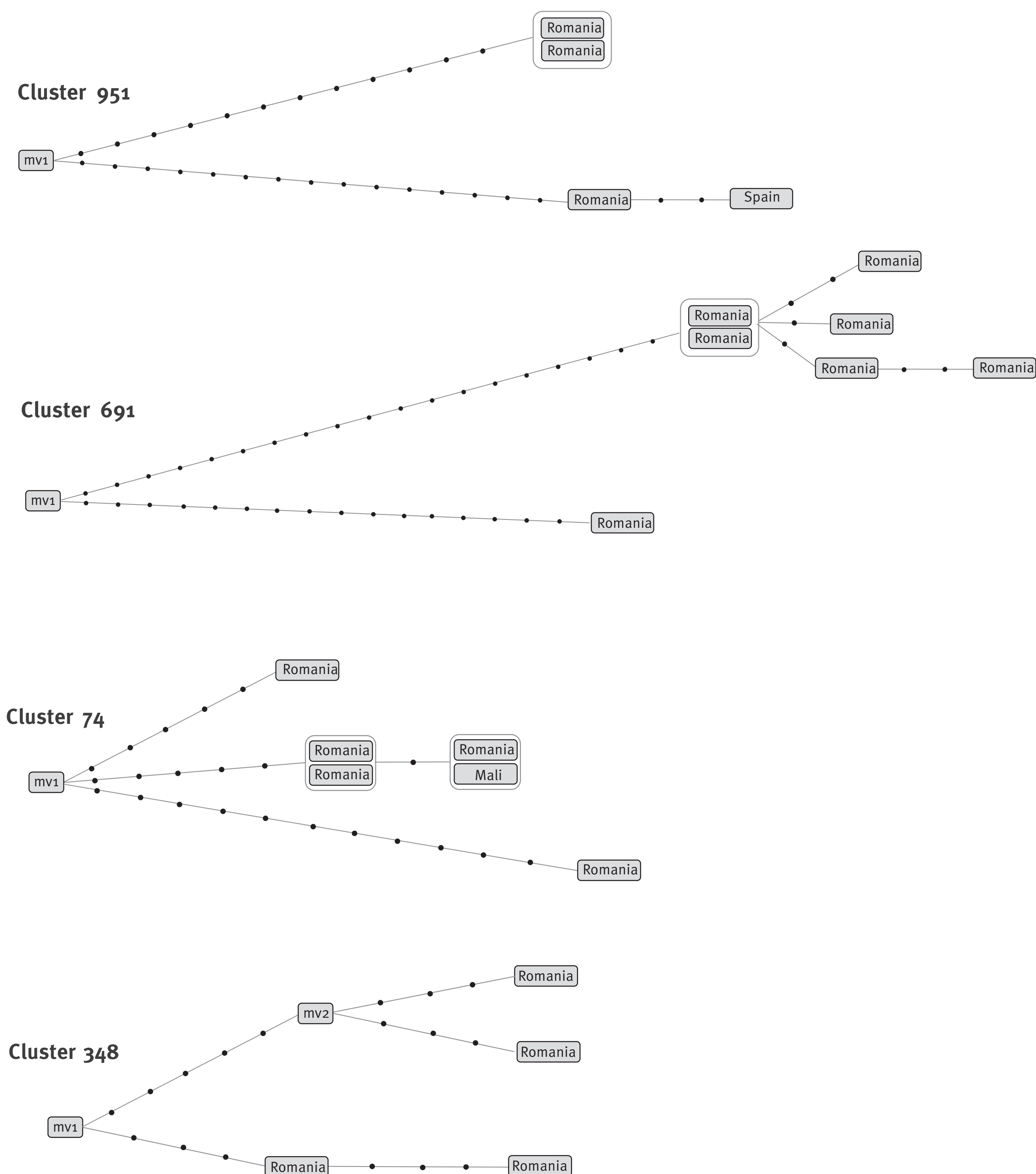
Two different ASO-PCRs were designed to analyse strain 113. The first ASO-PCR aimed to differentiate new secondary transmitted cases in Almería from independently imported cases. We designed a four-plex single-tube format. Two of the four SNPs targeted were strain 113-marker-SNPs (one targeted the 113 allele and the other the non-113 allele). The remaining two SNPs targeted were only shared by the 113-strain isolates involved in the recent transmission cluster (Supplementary Table S1). The design pursued to obtain three different amplification patterns depending on whether a new case corresponded to recent transmission by strain 113, importation of strain 113 or infection with a strain other than 113.

The reaction conditions were as follows: 1.5 mM $MgCl_2$, 0.2 µM of each primer (Supplementary Table S1), 200 µM deoxynucleotides (dNTPs) (Roche, Mannheim, Germany), 1% Dimethyl sulfoxide (DMSO) and 1.5 µL Taq DNA Polymerase (Roche, Mannheim, Germany). The PCR conditions were 95 °C for 5 min followed by 25–40 cycles (95 °C for 1 min, 61 °C for 1 min, and 72 °C for 1 min) and 72 °C for 10 min. The number of cycles was 25 when using as a template DNA purified from primary positive cultures and 40 when it was purified from sputa.

The second ASO-PCR was applied to assess whether an MTB isolate corresponded to strain 113 or to any other strain. We prepared another version of a four-plex single-tube ASO-PCR to target four SNPs (two alleles specific for isolates 113 and the other two alleles expected for non-113 strains) (Supplementary Table S2). Two different amplification patterns indicated whether a strain corresponded to the 113 strain or to any strain

Figure 3

Networks of relationships obtained from the whole genome sequencing analysis for clusters rich in cases from eastern Europe



mv: median vectors; SNP: single nucleotide polymorphisms.

Each black dot corresponds to a SNP. Each box corresponds to a patient and the nationality is indicated within the box. When two or more cases share identical sequences (zero SNPs between them) they are surrounded by a line.

other than 113. The reaction conditions were as follows: 1.5 mM $MgCl_2$, 0.2 µM of each primer (Supplementary Table S2), 200 µM dNTPs (Roche, Mannheim, Germany) and 1.5 µL Taq DNA Polymerase (Roche, Mannheim, Germany). The PCR conditions were 95 °C for 5 min followed by 30 cycles (95 °C for 1 min, 64 °C for 1 min, and 72 °C for 1 min) and 72 °C for 10 min. The ASO-PCR was applied on purified DNA purified or directly on bacterial lysates obtained from boiling stored frozen isolates.

The amplification patterns were analysed by sizing the amplification products using agarose gel electrophoresis.

Results

We selected 10 MIRU-VNTR–defined clusters (Figure 1) from the universal molecular epidemiology survey that has been running in Almería since 2003. The clusters were rich in cases from countries representative of three wide geographic areas, namely, sub-Saharan Africa (two clusters, in which most cases were from Senegal and Mali), north Africa (four clusters in which most cases were from Morocco) and eastern Europe (four clusters in which most cases were from Romania). All the involved strains were pansusceptible and corresponded to lineage four.

Sub-Saharan clusters

In cluster 1202, the analysis of SNPs from the 10 cases indicated the coexistence of a group of nine cases with a genetic distance of 0–7 SNPs between cases (Figure 2). The group included seven cases from Senegal, one from Morocco and one from Spain. Both observations strongly suggested that these nine cases were in fact part of a recent transmission event in Spain. Despite sharing an identical MIRU-VNTR pattern, the remaining case from Senegal showed a higher genetic distance i.e. 12 SNPs, with seven specific for this isolate and not sharing the five SNPs shared by all the isolates in the recent transmission group (Figure 2). These observations made it more likely, that this case corresponded to an unrelated importation from Senegal.

In cluster 789 (Figure 2), we sequenced five of the cases (four from Mali and the only case from Nigeria). The genetic distances between cases were 0–6 SNPs. No cases showed a distribution of SNPs that differed markedly within the group, suggesting the absence of independent importations from the country of origin.

Eastern European clusters

In three of the four clusters that were rich in cases from Romania (Figure 3), we detected the coincidence of cases due to either recent transmission or to independent importations.

In cluster 951, of the five cases, clustered by MIRU-VNTR, (Figure 3) WGS analysis of the four available isolates suggested that the theoretical cluster was hiding two independent subclusters. Two Romanian cases from the year 2011 differed in 27 SNPs and therefore corresponded to independent importations. Each case caused a secondary case in 2014 due to recent transmission in the host country. The isolates from the secondary cases had two SNPs (Spanish case) and zero SNPs (Romanian case) with respect to the corresponding index case.

A similar situation was observed for cluster 691 (Figure 3). WGS revealed that the MIRU-VNTR–defined cluster included two cases that brought together a high number of SNPs between them (35 SNPs), likely corresponding to two independent importations. A true recent transmission cluster had developed from one of these cases, with another five secondary cases occurring with genetic distances between cases of 0–5 SNPs. The other imported case corresponded to a dead-end branch i.e. it resulted in no secondary cases.

For cluster 74, we identified two different patterns (Figure 3). First, there were four highly related isolates, with 0–1 SNPs between cases, clearly indicative of recent transmission. Second, there were two branches, possibly corresponding to two independently imported cases with five and eleven specific SNPs, respectively, and did not share the five SNPs found in the four isolates belonging to the transmission subgroup. The transmission event (years 2003–2008) was caused by one of these likely imported cases, whereas the remaining two were representative of dead-end branches (years of isolation: 2013 and 2015).

Finally, in cluster 348 (Figure 3), two cases had a genetic distance of three SNPs, suggesting recent transmission between them. However, a definitive interpretation could not be found for the remaining two cases. The cases showed a genetic distance of six SNPs between them, but a non-sampled node (mv2) was inferred to be located between them in the network. It is, therefore, unclear whether these two cases are part of a recent transmission chain involving a non-sampled case in Spain or if they corresponded to two imported cases that were epidemiologically related with a non-sampled case at the host country.

North African clusters

In three of the four clusters, predominately comprising of cases from Morocco, short genetic distances were recorded between all clustered cases (cluster 558: 0–5 SNPs, cluster 1192: 0–3 SNPs and cluster 5: 0–2 SNPs between cases), highly indicative of recent transmission in the host country, Spain (Figure 4).

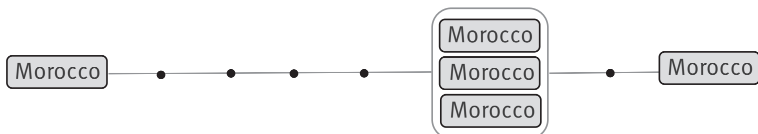
However, for the remaining cluster, cluster 113, which included 17 cases, WGS of the 14 available isolates revealed a much more complex network of relationships (Figure 4).

Three median vectors (mv) corresponding to non-sampled cases had to be defined. Seven independent branches were observed (Figure 4), with four, four, seven, eight, nine, 10 and 13 specific SNPs for each of

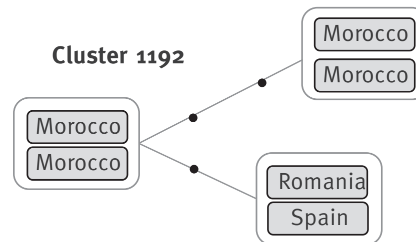
Figure 4
Networks of relationships obtained from the whole genome sequencing analysis for clusters rich in cases from North Africa

North Africa

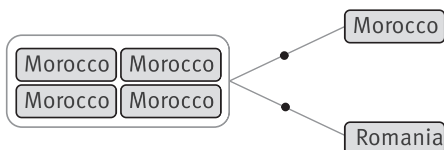
Cluster 558



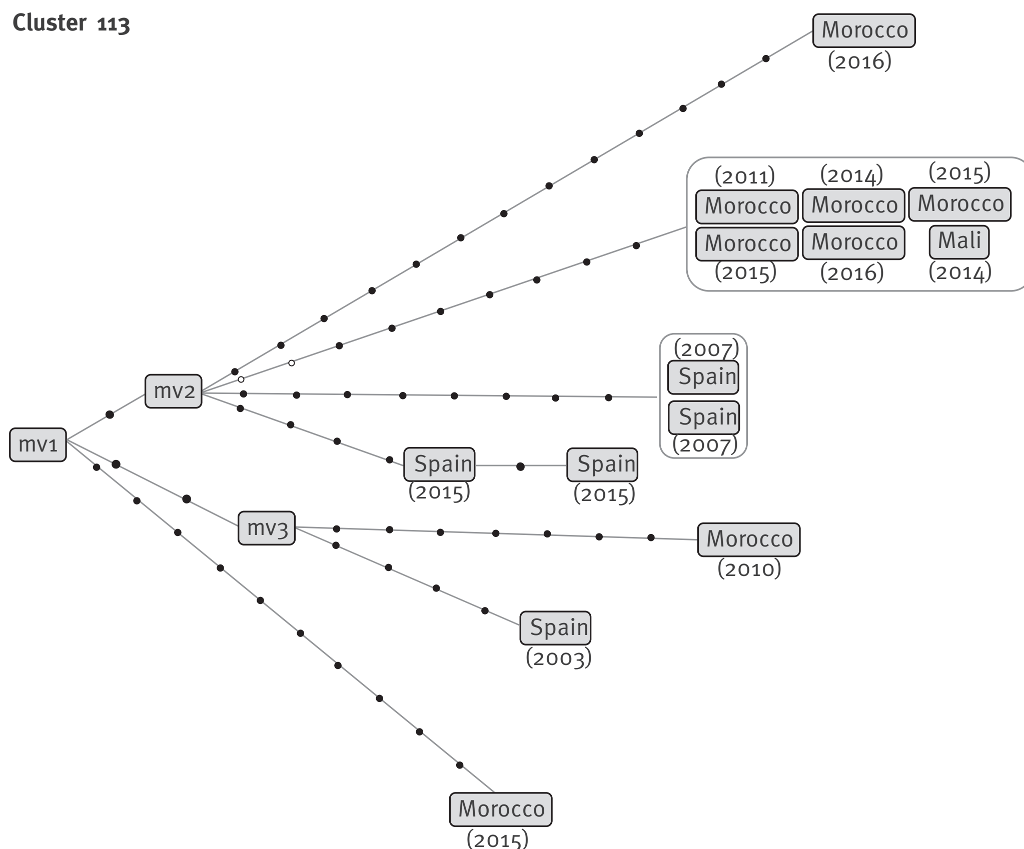
Cluster 1192



Cluster 005



Cluster 113

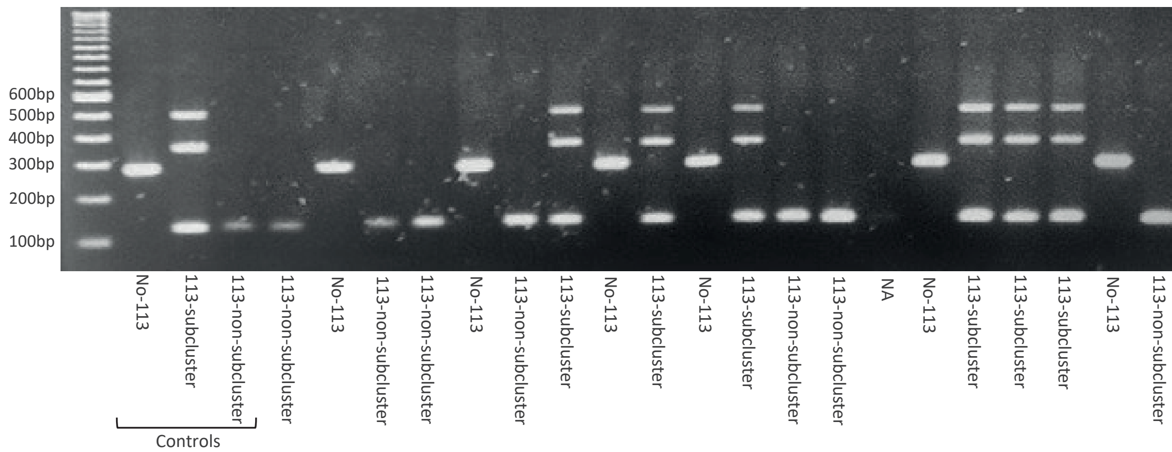


mv: median vectors. SNP: single nucleotide polymorphisms.

Each box corresponds to a patient and the nationality is indicated within the box. When two or more cases share identical sequences (0 SNPs between them) they are surrounded by a line. Each black dot corresponds to a SNP. White dots detailed in cluster 113 correspond to non-fixed alleles, found in heterozygosis in one of the cases, but in homozygosis in the remaining cases. The years of diagnosis are indicated in brackets.

Figure 5

Results for the multiplex ASO-PCR designed to precisely assign new incident cases infected by the strain 113 in Almería and labelling them as due to recent transmission or importation.



Amplification patterns obtained from a selection of isolates representative of the 113 transmission subcluster (113 subcluster), isolates 113 not included in the recent transmission subcluster (113 non-subcluster) and isolates other than 113 (No 113). The different amplification patterns for each group can be observed.

the branches and each more likely corresponding to unrelated cases (distances between each two branches were in the range of 11–24 SNPs). Therefore, these cases were likely due to unrelated importations from Morocco. Of the seven branches, four corresponded to dead-ends, including a single case each (years 2003, 2010, 2015, and 2016); three were from Morocco and were diagnosed 10, 6, and 2 years after arrival. As there were no additional related secondary cases, the findings seem consistent with likely reactivations.

Two of the remaining three branches showed one additional case that was closely related to the imported index case in each branch (zero and one SNPs), which was diagnosed the same year as the index case (year 2007 and 2015, respectively), possibly due to self-limited recent transmission events in Spain.

The remaining branch was the only one with a higher number of cases i.e. six, among which no SNPs were found. Of note, two alleles were in heterozygosis in one of these cases (year 2011) and were fixed as homozygotes in the remaining five cases. Based on this observation, we can infer that the case with heterozygosis was the index case and the remaining five cases were secondary cases and likely due to recent transmissions in Spain.

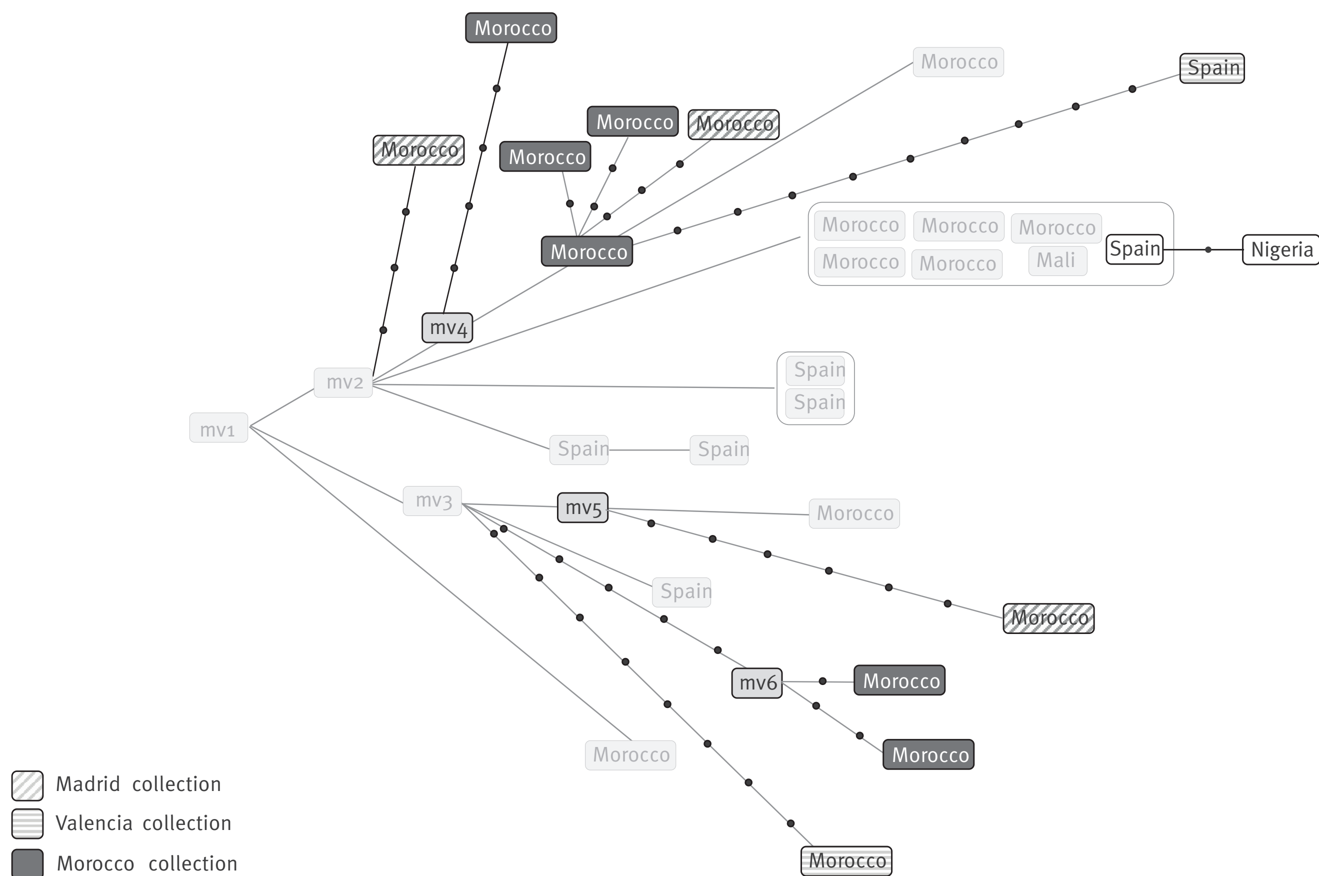
New strategy based on whole genome sequencing data to precisely identify recent transmission

In our context, MIRU-VNTR was proved useless, because it could not discriminate between the three events observed for strain 113 e.g. dead end-imported hosts, self-limited transmission chains and ongoing active transmission events. Among the 17 cases theoretically linked by MIRU-VNTR, only six were really involved in an active recent transmission chain whereas the remaining 11 cases had been misclassified and their epidemiological follow-up was not well oriented. Using standardised interviews with the cases it was possible to establish epidemiological links between the cases in the six-case subcluster, revealing that three cases were customers of the same bar and another case shared a flat with them.

In order to be able to precisely identify the true secondary cases in an active transmission chain, we defined a new approach. We first identified the 71 common SNPs shared by all members in MIRU-VNTR–defined cluster 113 and those SNPs which were specific for the different branches in the network. We designed an allele-specific multiplex PCR (ASO-PCR) including four PCRs, which targeted the following (Supplementary Table S1): (i) two SNPs specific for all the strain 113 isolates in the network, which were selected as a general marker for this strain (one PCR targeting the 113 allele in one

Figure 6

Extended network of relationships obtained from the whole genome sequencing analysis for cluster 113 including Almería, Madrid, Valencia and Morocco isolates



mv: median vectors; SNP: single nucleotide polymorphisms.

Each box corresponds to a patient and the nationality is indicated within the box. When two or more cases share identical sequences (0 SNPs between them) they are surrounded by a line. Each black dot corresponds to a SNP. Isolates in Almería detailed in Figure 4 are now fainted. The two new cases identified in Almería by applying the ASO-PCR are shown in white boxes.

of these SNPs and the other PCR the non-113 allele from the other SNP) and (ii) two SNPs among the nine SNPs that were only shared by the branch including the active transmission subcluster (targeting the alleles for the active transmission subcluster).

The ASO-PCR was designed following a four-plex format to target the four SNPs simultaneously in the same reaction tube. This lead to three different amplification patterns depending on whether a new case corresponded to the recently transmitted subcluster 113, to a 113 isolate not involved in this active subcluster (therefore corresponding to a new unrelated importation) or to a strain other than 113 (Figure 5). The specificity of the multiplex ASO-PCR was checked by testing all the 14 isolates with the 113 VNTR pattern and a selection of 20 randomly selected strains for Moroccan migrants among those diagnosed in Almería. The expected pattern for the three possible profiles was obtained in all cases.

The PCR was transferred to Torrecardenas Hospital in Almería to be prospectively applied on all newly diagnosed TB cases of Moroccan origin or living in the same area as the cases involved in the MIRU-VNTR–defined cluster 113. We first checked that the PCR was sensitive enough to be applied directly on respiratory specimens and were able to obtain an interpretable profile when decontaminated sputa with high or medium bacillary load were used as templates.

An interpretable result was obtained for all the eight stain-positive cases in which the multiplex ASO-PCR was prospectively applied (during a 3-month period) directly on sputa. For the prospective cases with paucibacillari sputa it was necessary to wait until culture was available. In 15 cases, the pattern corresponded

to a non-113 strain; however, in two cases (one from Spain in 2016 and the other from Nigeria in 2017), we obtained the pattern expected for active subcluster 113. Both isolates shared the expected 113 MIRU-VNTR pattern. Subsequent WGS analysis indicated that they showed zero and one SNPs with the six isolates previously included in the active subcluster (Figure 6).

Expanded analysis of strain 113 in unrelated populations

Once the demand for identification of new cases due to recent transmission of the active transmission node was resolved, we focused on the other issue affecting this cluster i.e. the independent importations of closely related (genetically) strains from the country of origin, those likely prevalent in Morocco that have acquired diversity by circulating over extended periods of time. We tried to identify other examples of independent importations for this strain in other unrelated populations.

For this purpose, we selected two Spanish populations: one from Valencia (eastern Spain), a representative of a node with WGS data available from a population-based genomic epidemiology programme and another one from Madrid (central Spain), for which no population-based WGS data were available.

The approach in Valencia was direct and limited to querying on the presence of the 71 SNPs that are specific for the isolates in cluster 113; we identified two cases sharing all the 71 SNPs. When these were integrated into the Almería network, they consistently corresponded to two new subbranches in two of the previously described importation branches (Figure 6).

The approach in Madrid was indirect, involving application of a multiplex ASO-PCR directly on stored isolates from Moroccan migrant TB cases. We prepared a new version of a four-plex ASO-PCR to target four SNPs. Two of the PCRs targeted the alleles that were specific for isolates 113 and the other two targeted the alleles expected for non-113 strains (Supplementary Table S2); the two amplification patterns identified indicated whether a strain corresponded to the 113 MIRU-VNTR cluster or to any strain other than 113 (Figure 7). We applied it to 134 available Moroccan isolates from our retrospective convenience sample and detected the 113 pattern in five cases (Figure 7a). WGS of three of these isolates confirmed them to be 113 (they included all 71 SNPs) and their integration in the network revealed three new branches (Figure 6).

Expanded analysis of strain 113 in the country of origin

We completed the general analysis of strain 113, with a cross-border analysis, by tracking its circulation in the country of origin. The epidemiological information collected from cases by interview aided in determining that most migrant cases were from cities in the north of Morocco.

Molecular epidemiology studies in northern Morocco were checked in which MIRU-VNTR genotypes corresponding to strain 113 could be found. Chaoui et al. [23] reported a cluster involving four cases in Tangier infected by a LAM3 SIT33 strain that could correspond to strain 113. However, only data for the 12-loci version of MIRU-VNTR were available.

To confirm whether strain 113 was circulating in the area, as suggested by the published data, the same multiplex ASO-PCR that had been designed to track strain 113 in Madrid was transferred and locally applied in Morocco. Interrogation of 11 SIT33 isolates revealed seven with the pattern corresponding to strain 113. In addition, testing of 45 additional retrospective isolates from northern Morocco (Tangier, Tetouan and Larache), for which no previous genotypic information was available, revealed a 113 pattern in seven isolates (Figure 7b). WGS was performed in six of the 14 isolates that were positive for 113 and enabled us to integrate them into the network of relationships (Figure 6). Three of the isolates were positioned in two new sub-branches and the other three were located in one of a previously defined importation branch. Furthermore, two probable recent transmission events in Morocco, involving two and three cases respectively, were identified indirectly (with three SNPs between cases in both of them).

Discussion

Molecular epidemiology based on universal genotyping of TB cases in a population allows us to identify clustered cases that are infected by *M. tuberculosis* isolates with identical fingerprints. From the analysis of clustered cases, we can obtain valuable data on transmission dynamics in different epidemiological scenarios.

The increased complexity resulting from changing socio-epidemiological features due to migration demands special attention. The clusters may be autochtonous, mixed multinational, and foreign-born clusters rich in cases from a specific country.

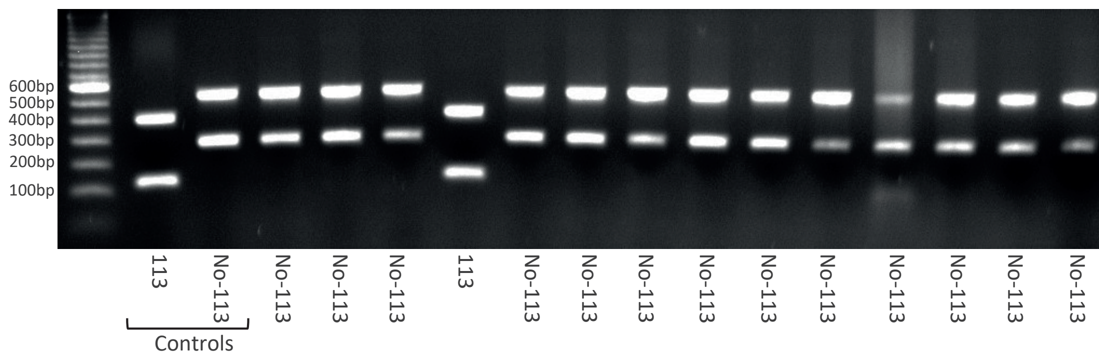
Some of the complex molecular clusters identified in populations with a higher percentage of migrants are not always accompanied by clear epidemiological links between the cases involved [7,24,25]. Here, we tried to analyse whether the lack of epidemiological support could mean that some of the clusters involving migrants were not robust and were misleadingly alerting us to recent transmissions.

We hypothesised that some of the cases in these clusters could correspond to independent importations of strains that might be prevalent in the country of origin. Genetic diversity would be expected to accumulate for a prevalent strain circulating in a high-incidence country over extended periods of time. However, the diversity accumulated is probably insufficient to lead to a change in the MIRU-VNTR pattern, thus explaining why unrelated cases independently importing these strains

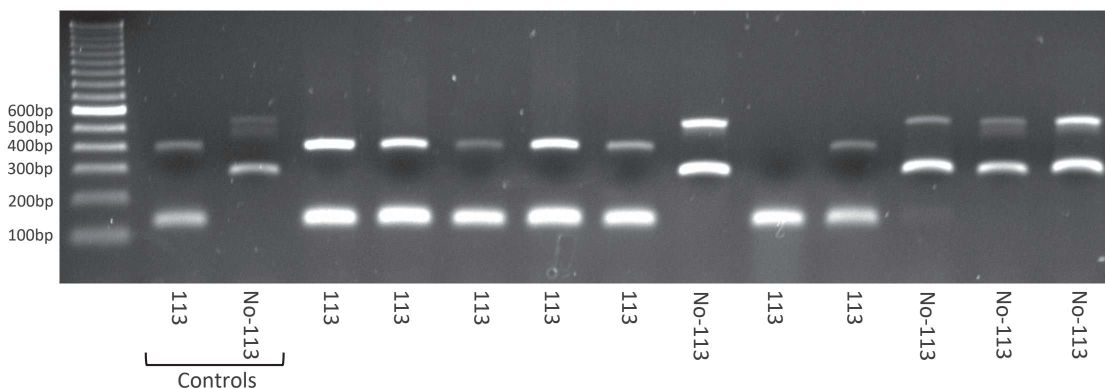
Figure 7

Results for the multiplex ASO-PCR designed to (A) retrospectively track strain 113 in Madrid and (B) retrospectively track strain 113 in Morocco

A. Multiplex ASO-PCR results, retrospectively track strain 113, Madrid



B. Multiplex ASO-PCR results, retrospectively track strain 113, Morocco



The different amplification patterns for the strains belonging to the 113 group and those that do not are observed.

appear clustered. MIRU-VNTR types are conserved for highly prevalent strains, as reported in Denmark for a highly prevalent strain responsible for 35% cases over 15 years [26]. However, the application of more discriminative methods e.g. WGS, could help us to reveal some degree of diversity between these prevalent strains and differentiate between true recent transmission in the host country (when no or very limited genetic diversity is found between the corresponding isolates) and independent, unrelated importations of prevalent strains in the country of origin (if we detect greater genetic distance).

Application of this strategy, following the consensus thresholds of diversity to assign or rule out recent transmission with WGS [12], revealed that unrelated importations were hidden within some MIRU-VNTR–defined clusters and had been misinterpreted as recent transmissions in the host country. Due to the size of certain clusters in the analysis we only revealed a minority (one case in several clusters) that had been misclassified as recent transmission when it was really due to importation. However, in some of the bigger clusters, the magnitude of misclassified cases revealed was higher (eight of 14).

In a 2016 publication, Stucki et al. [27] reported importations within MIRU-VNTR clusters in a nationwide analysis in Switzerland (90 patients in 35 clusters during 2000–08). Only 25% of the MIRU-VNTR–defined clusters including migrants (in this case, mostly from east Africa) were refuted using WGS. The clustering proportion fell from 16.7% to 6.5% for migrant clusters; when only Swiss-born clusters were considered, the decrease was smaller (19.3% to 14.3%). In addition, descriptions of misassignation of recent transmission in MIRU-VNTR–proven migrant clusters revealed by WGS have recently been reported in Canada [28] and the Netherlands [29].

Although our findings are limited to the low number of clusters selected, both these data and ours suggest that the involvement of genetically closely related strains imported independently from high-incidence regions is a widespread phenomenon. We cannot extend the findings from the migrant clusters in our study to all clusters including migrants because in our setting some migrant nationalities were not represented. Nevertheless, our results showed that this phenomenon was not anecdotal or restricted to specific geographic areas and that it was found in clusters with migrants that were representative of different areas e.g. eastern Europe, north Africa and sub-Saharan Africa.

In our study, the identification of imported cases within clusters defined by standard genotyping was mainly supported by the analysis of the total number of SNPs between the clustered cases. However, the analysis of the chronology of diagnosis of the TB cases can also be useful to identify importations. This is because the order of emergence of SNPs is sequential and once acquired they do not reverse [30]. In cluster 558, the last case diagnosed (year 2014) did not present the four SNPs identified in the remaining clustered cases, diagnosed 3–8 years earlier. The most likely explanation is that the 2014 case was imported from a more ancestral branch than the one involved in the recent transmission event in Spain.

The demonstration that both importations and recent transmissions could co-exist in a cluster defined by standard genotyping raised an alert: once one of these genetically closely related strains is imported into a host country, standard molecular epidemiology–surveillance approaches are of very limited value. Based on standard MIRU-VNTR, it would be impossible to discriminate between secondary cases that originated in the host country and unrelated independent importations: all cases would be equally considered clustered.

It is important to differentiate between a new imported case and a recently transmitted secondary case, because each represents a completely different epidemiological situation that has to be managed separately. Consequently, other authors have recommended WGS as the only way to ensure more accurate identification of recent transmission, particularly among migrants from high-incidence areas [27,31]. An alternative to the analysis based on WGS and SNPs calling based on pipelines is the technique of core genome MLST typing, which takes advantage to the discriminatory power of the next generation sequencing (NGS) technique and makes easier the SNP calling by standardised processing and allows a more direct comparative analysis across different laboratories [32]. However, global implementation of WGS is expensive and WGS has been successfully implemented at population level in few settings only [33-35]. With the aim to overcome these limitations and to find a solution that can be implemented in settings where nationwide WGS application is not a reality, we adapted a strategy previously developed by our group to survey high-risk strains. This strategy is based on tailored ASO-PCRs targeting strain-specific SNPs identified from WGS data of representative isolates for the strains to be surveyed [36]. We implemented it in previous studies to be able to provide a fast response to challenges, such as optimising surveillance of transmission of actively transmitted strains [36], rapid tracking of the presence of specific outbreak strains in a population [37] and confirming the presence of secondary cases due to imported XDR strains from Russia directly on respiratory specimens in the hospital setting [38]. In the current study we adapted the strategy to tailor PCRs targeting the SNPs that were specific for isolates actively involved in recent transmission in the host country and to differentiate these isolates from other independently imported isolates which lacked those SNPs.

To pilot this strategy, we selected the most complex cluster in our study, namely cluster 113, which was rich in cases from Morocco (six different importation branches together with an active transmission cluster). The strategy prospectively identified new secondary cases directly from respiratory specimens. Our proposal not only resolved the epidemiological challenge at the local level, but also enabled us to expand the boundaries of our analysis to other unrelated populations in Spain. If this strain corresponded to a prevalent strain in the country of origin, we would be able to find it in unrelated populations receiving migrants from Morocco. We identified the strain in the two unrelated populations surveyed and proved that importations of the same strain occurred in other settings, thus showing that they were not the result of recent transmissions. For some of the remaining studied strains from migrants from Morocco we also found data indicating they are circulating also in Morocco [23,39] and similar efforts could be done to fully characterise their global distribution.

Conclusion

Tracking transmission of TB through cross-border surveillance is a crucial element in the current epidemiological surveillance of TB, and data from both the country of origin and host countries must be integrated as recently exemplified in a study revealing a cross-border MDR-TB cluster involving several European countries [40]. Our findings revealed standard MIRU-VNTR-based epidemiology was not a suitable approach for cross-border surveillance as it was unable to discriminate between importations and recent transmissions. WGS-based analysis was able to differentiate these two overlapping events, however, genomic analysis is not accessible for many countries involved in cross-border TB transmission. Here, we propose a new strategy, adapted to settings with no or limited access to WGS , based on designing simple PCR tools tailored to be adapted to identify either recent transmission in the host country or independent importations from the

country of origin. Adapted versions of the same PCRs were also designed to be transferred and applied to track the strain circulating in the country of origin.

Our next step will be to extend the approaches used in this study to develop a network of nodes surveying prevalent strains from countries with a high TB incidence that are being exported to countries with low-TB burden. Such a network could contribute to the establishment of a new global cross-border surveillance system, fitted to the challenges associated with international migration.

Acknowledgements

We thank Thomas O'Boyle for proofreading the manuscript. We also thank Remedios Guna (Hospital General de Valencia), Lina Gimeno (Hospital General de Alicante) and Isabel Escribano (Hospital Virgen de los Lirios, Alcoy) for providing the clinical isolates from the Comunidad Valenciana analyzed in this study. This project was funded by ISCIII: ERANET-LAC (TRANS-TB-TRANS REF AC16/00057; ELAC2015/T08-0664), FIS (13/01207; 15/01554) and cofunded by ERDF Funds from the European Commission: "A way of making Europe". Miguel Servet grant (CP15/00075) for LPL. Ministerio de Economía y Competitividad (grant SAF2016-77346-R), ERC (638553-TB-ACCELERATE) to IC. FPU13/00913 (Ministerio de Educación y Ciencia) to ACO.

Conflict of interest

None declared.

Authors' contributions

Design of the study, analysis of results, preparation of the MS: DGV, LPL, EA Experimental tasks: EA, LPL, MML, MH, IC, ICH, JAGC, SS. Analysis of data: DGV, EA, LPL, MML, ACO, MH, IC, MDEM. General revision: EB.

References

1. de Vries G, Aldridge RW, Cayla JA, Haas WH, Sandgren A, van Hest NA, et al. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. Euro Surveill. 2014;19(9):20726. https://doi.org/10.2807/1560-7917.ES2014.19.9.20726 PMID: 24626208
2. Diel R, Rüsch-Gerdes S, Niemann S. Molecular epidemiology of tuberculosis among immigrants in Hamburg, Germany. J Clin Microbiol. 2004;42(7):2952-60. https://doi.org/10.1128/JCM.42.7.2952-2960.2004 PMID: 15243044
3. Iñigo J, García de Viedma D, Arce A, Palenque E, Herranz M, Rodríguez E, et al. Differential findings regarding molecular epidemiology of tuberculosis between two consecutive periods in the context of steady increase of immigration. Clin Microbiol Infect. 2013;19(3):292-7. https://doi.org/10.1111/j.1469-0691.2012.03794.x PMID: 22404140
4. Ospina JE, Orcau À, Millet JP, Ros M, Gil S, Caylà JABarcelona Tuberculosis Immigration Working Group. Epidemiology of Tuberculosis in Immigrants in a Large City with Large-Scale Immigration (1991-2013). PLoS One. 2016;11(10):e0164736. https://doi.org/10.1371/journal.pone.0164736 PMID: 27749904
5. De Beer JL, Kodmon C, van der Werf MJ, van Ingen J, van Soolingen D, Collective the ECDC MDR-TB molecular surveillance project participants. Molecular surveillance of multi- and extensively drug-resistant tuberculosis transmission in the European Union from 2003 to 2011. Euro Surveill. 2014;19(11):20742. https://doi.org/10.2807/1560-7917.ES2014.19.11.20742 PMID: 24679719
6. Lillebaek T, Andersen AB, Bauer J, Dirksen A, Glismann S, de Haas P, et al. Risk of Mycobacterium tuberculosis transmission in a low-incidence country due to immigration from high-incidence areas. J Clin Microbiol. 2001;39(3):855-61. https://doi.org/10.1128/JCM.39.3.855-861.2001 PMID: 11230395
7. Martínez-Lirola M, Alonso-Rodriguez N, Sánchez ML, Herranz M, Andrés S, Peñafiel T, et al. Advanced survey of tuberculosis transmission in a complex socioepidemiologic scenario with a high proportion of cases in immigrants. Clin Infect Dis. 2008;47(1):8-14. https://doi.org/10.1086/588785 PMID: 18484876
8. Alonso Rodríguez N, Chaves F, Iñigo J, Bouza E, García de Viedma D, Andrés S, et al. TB Molecular Epidemiology Study Group of Madrid. Transmission permeability of tuberculosis involving immigrants, revealed by a multicentre analysis of clusters. Clin Microbiol Infect. 2009;15(5):435-42. https://doi.org/10.1111/j.1469-0691.2008.02670.x PMID: 19416291
9. Bryant JM, Schürch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, et al. Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. BMC Infect Dis. 2013;13(1):110. https://doi.org/10.1186/1471-2334-13-110 PMID: 23446317
10. 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. https://doi.org/10.1056/NEJMoa1003176 PMID: 21345102
11. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, et al. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med. 2013;10(2):e1001387. https://doi.org/10.1371/journal.pmed.1001387 PMID: 23424287
12. Walker TM, Ip CL, 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. 2013;13(2):137-46. https://doi.org/10.1016/S1473-3099(12)70277-3 PMID: 23158499
13. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, Bonillo MC, et al. Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for fast identification of clustered and orphan cases. J Clin Microbiol. 2009;47(7):2026-32. https://doi.org/10.1128/JCM.02308-08 PMID: 19458183
14. 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. https://doi.org/10.1128/JCM.01392-06 PMID: 17005759
15. Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, et al. Two new rapid SNP-typing methods for classifying Mycobacterium tuberculosis complex into the main phylogenetic lineages. PLoS One. 2012;7(7):e41253. https://doi.org/10.1371/journal.pone.0041253 PMID: 22911768
16. Carcelén M, Abascal E, Herranz M, Santantón S, Zenteno R, Ruiz Serrano MJ, et al. Optimizing and accelerating the assignation of lineages in Mycobacterium tuberculosis using novel alternative single-tube assays. PLoS One. 2017;12(11):e0186956. https://doi.org/10.1371/journal.pone.0186956 PMID: 29091913
17. Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, et al. Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. J Infect Dis. 2014;209(1):98-108. https://doi.org/10.1093/infdis/jit439 PMID: 23945373
18. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013;45(10):1176-82. https://doi.org/10.1038/ng.2744 PMID: 23995134
19. The R project for Statistical computing. Available from: https://www.r-project.org
20. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014;5(1):4812. https://doi.org/10.1038/ncomms5812 PMID: 25176035
21. Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, et al. Population Genomics of Mycobacterium tuberculosis in Ethiopia Contradicts the Virgin Soil Hypothesis for Human Tuberculosis in Sub-Saharan Africa. Curr Biol. 2015;25(24):3260-6. https://doi.org/10.1016/j.cub.2015.10.061 PMID: 26687624

22. Guerra-Assunção JA, Crampin AC, Houben RM, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. eLife. 2015;4:e05166. https://doi.org/10.7554/eLife.05166 PMID: 25732036
23. Chaoui I, Zozio T, Lahlou O, Sabouni R, Abid M, El Aouad R, et al. Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent Mycobacterium tuberculosis genotypes infecting tuberculosis patients in Morocco. Infect Genet Evol. 2014;21:463-71. https://doi.org/10.1016/j.meegid.2013.05.023 PMID: 23732366
24. Anderson LF, Tamne S, Brown T, Watson JP, Mullarkey C, Zenner D, et al. Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. Lancet Infect Dis. 2014;14(5):406-15. https://doi.org/10.1016/S1473-3099(14)70022-2 PMID: 24602842
25. Pedersen MK, Andersen AB, Andersen PH, Svensson E, Jensen SG, Lillebaek T. Occupational Tuberculosis in Denmark through 21 Years Analysed by Nationwide Genotyping. PLoS One. 2016;11(4):e0153668. https://doi.org/10.1371/journal.pone.0153668 PMID: 27082745
26. Kamper-Jørgensen Z, Andersen AB, Kok-Jensen A, Bygbjerg IC, Andersen PH, Thomsen VO, et al. Clustered tuberculosis in a low-burden country: nationwide genotyping through 15 years. J Clin Microbiol. 2012;50(8):2660-7. https://doi.org/10.1128/JCM.06358-11 PMID: 22675129
27. Stucki D, Ballif M, Egger M, Furrer H, Altpeter E, Battegay M, et al. Standard Genotyping Overestimates Transmission of Mycobacterium tuberculosis among Immigrants in a Low-Incidence Country. J Clin Microbiol. 2016;54(7):1862-70. https://doi.org/10.1128/JCM.00126-16 PMID: 27194683
28. Guthrie J, Kong C, Roth D, Rodrigues M, Hoang L, Walker T, et al. Findings from whole genome sequencing of tuberculosis in a geographically large Canadian province with a diverse population. 25-28 Jun 2017. 38th Congress of the European Society of mycobacteriology.; Sibenik, Croatia. OP 29. Available from: http://vorschau.agentur-konsens.de/ESM/downloads/2017/ESM_Abstract_Book2017_final.pdf
29. Jajou R dNH, Mulder A, Kamst M, van Hunen R, de Vries G, Anthony R, et al. VNTR typing is less reliable in predicting epidemiological links than expected, as indicated by whole genome sequencing (WGS) analysis. J Clin Microbiol. Forthcoming 2017.
30. Walker TM, Monk P, Smith EG, Peto TE. Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect. 2013;19(9):796-802. https://doi.org/10.1111/1469-0691.12183 PMID: 23432709
31. Folkvardsen DB, Norman A, Andersen AB, Michael Rasmussen E, Jelsbak L, Lillebaek T. Genomic Epidemiology of a Major Mycobacterium tuberculosis Outbreak: Retrospective Cohort Study in a Low-Incidence Setting Using Sparse Time-Series Sampling. J Infect Dis. 2017;216(3):366-74. https://doi.org/10.1093/infdis/jix298 PMID: 28666374
32. Kohl TA, Diel R, Harmsen D, Rothgänger J, Walter KM, Merker M, et al. Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. J Clin Microbiol. 2014;52(7):2479-86. https://doi.org/10.1128/JCM.00567-14 PMID: 24789177
33. Cirillo DM, Cabibbe AM, De Filippo MR, Trovato A, Simonetti T, Rossolini GM, et al. Use of WGS in Mycobacterium tuberculosis routine diagnosis. Int J Mycobacteriol. 2016;5(Suppl 1):S252-3. https://doi.org/10.1016/j.ijmyco.2016.09.053 PMID: 28043590
34. Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. Lancet Respir Med. 2014;2(4):285-92. https://doi.org/10.1016/S2213-2600(14)70027-X PMID: 24717625
35. Walker TM, Merker M, Kohl TA, Crook DW, Niemann S, Peto TE. Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. Clin Microbiol Infect. 2017;23(3):161-6. https://doi.org/10.1016/j.cmi.2016.10.014 PMID: 27789378
36. Perez-Lago L, Martinez Lirola M, Herranz M, Comas I, Bouza E, Garcia-de-Viedma D. Fast and low-cost decentralized surveillance of transmission of tuberculosis based on strain-specific PCRs tailored from whole genome sequencing data: a pilot study. Clin Microbiol Infect. 2015;21(3):249 e1-9.
37. Pérez-Lago L, Herranz M, Comas I, Ruiz-Serrano MJ, López Roa P, Bouza E, et al. Ultrafast Assessment of the Presence of a High-Risk Mycobacterium tuberculosis Strain in a Population. J Clin Microbiol. 2016;54(3):779-81. https://doi.org/10.1128/JCM.02851-15 PMID: 26719445
38. Pérez-Lago L, Martínez-Lirola M, García S, Herranz M, Mokrousov I, Comas I, et al. Urgent Implementation in a Hospital Setting of a Strategy To Rule Out Secondary Cases Caused by Imported Extensively Drug-Resistant Mycobacterium tuberculosis Strains at Diagnosis. J Clin Microbiol. 2016;54(12):2969-74. https://doi.org/10.1128/JCM.01718-16 PMID: 27682128
39. Lahlou O, Millet J, Chaoui I, Sabouni R, Filali-Maltouf A, Akrim M, et al. The genotypic population structure of Mycobacterium tuberculosis complex from Moroccan patients reveals a predominance of Euro-American lineages. PLoS One. 2012;7(10):e47113. https://doi.org/10.1371/journal.pone.0047113 PMID: 23077552
40. Fiebig L, Kohl TA, Popovici O, Mühlenfeld M, Indra A, Homorodean D, et al. A joint cross-border investigation of a cluster of multidrug-resistant tuberculosis in Austria, Romania and Germany in 2014 using classic, genotyping and whole genome sequencing methods: lessons learnt. Euro Surveill. 2017;22(2):30439. https://doi.org/10.2807/1560-7917.ES.2017.22.2.30439 PMID: 28106529

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Cocirculation of Hajj and non-Hajj strains among serogroup W meningococci in Italy, 2000 to 2016

Cecilia Fazio[1], Arianna Neri[1], Paola Vacca[1], Andrea Ciammaruconi[2], Milena Arghittu[3], Anna Maria Barbui[4], Caterina Vocale[5], Paola Bernaschi[6], Patrizia Isola[7], Irene Alessandra Galanti[8], Antonella Mencacci[9], Rosella De Nittis[10], Maria Chironna[11], Anna Giammanco[12], Elisabetta Pagani[13], Alessandro Bisbano[14], Paola Stefanelli[1]

1. Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy
2. Molecular Biology Section, Army Medical and Veterinary Research Center, Rome, Italy
3. Microbiology Unit, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy
4. Microbiology and Virology Laboratory, Molinette Hospital, Turin, Italy
5. Unit of Clinical Microbiology, Regional Reference Centre for Microbiological Emergencies, St. Orsola Malpighi University Hospital, Bologna, Italy
6. Microbiology Laboratory, Bambino Gesù Hospital, Rome, Italy
7. Clinical Pathology Department, Azienda USL 6, Livorno, Italy
8. Microbiology Laboratory, Azienda USL Toscana sud est, Arezzo, Italy
9. Medical Microbiology Section, Dept. of Medicine, University of Perugia, Perugia, Italy
10. Clinical Pathology Department, University Hospital, Foggia, Italy
11. Biomedical Sciences and Human Oncology Department – Hygiene Section, University Hospital, Bari, Italy
12. Department of Sciences for Health Promotion and Mother and Child Care "G. D'Alessandro", University of Palermo, Palermo, Italy
13. Microbiology and Virology Laboratory, Azienda Sanitaria dell'Alto Adige, Bolzano, Italy
14. Epidemiology Unit ASP Crotone, Italy

Correspondence: Paola Stefanelli (paola.stefanelli@iss.it)

Citation style for this article:
Fazio Cecilia, Neri Arianna, Vacca Paola, Ciammaruconi Andrea, Arghittu Milena, Barbui Anna Maria, Vocale Caterina, Bernaschi Paola, Isola Patrizia, Galanti Irene Alessandra, Mencacci Antonella, De Nittis Rosella, Chironna Maria, Giammanco Anna, Pagani Elisabetta, Bisbano Alessandro, Stefanelli Paola. Cocirculation of Hajj and non-Hajj strains among serogroup W meningococci in Italy, 2000 to 2016. Euro Surveill. 2019;24(4):pii=1800183. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800183

Article submitted on 11 Apr 2018 / accepted on 23 Oct 2018 / published on 24 Jan 2019

In Italy, B and C are the predominant serogroups among meningococci causing invasive diseases. Nevertheless, in the period from 2013 to 2016, an increase in serogroup W Neisseria meningitidis (MenW) was observed. This study intends to define the main characteristics of 63 MenW isolates responsible of invasive meningococcal disease (IMD) in Italy from 2000 to 2016. We performed whole genome sequencing on bacterial isolates or single gene sequencing on culture-negative samples to evaluate molecular heterogeneity. Our main finding was the cocirculation of the Hajj and the South American sublineages belonging to MenW/clonal complex (cc)11, which gradually surpassed the MenW/cc22 in Italy. All MenW/cc11 isolates were fully susceptible to cefotaxime, ceftriaxone, ciprofloxacin, penicillin G and rifampicin. We identified the full-length NadA protein variant 2/3, present in all the MenW/cc11. We also identified the fHbp variant 1, which we found exclusively in the MenW/cc11/Hajj sublineage. Concern about the epidemic potential of MenW/cc11 has increased worldwide since the year 2000. Continued surveillance, supported by genomic characterisation, allows high-resolution tracking of pathogen dissemination and the detection of epidemic-associated strains.

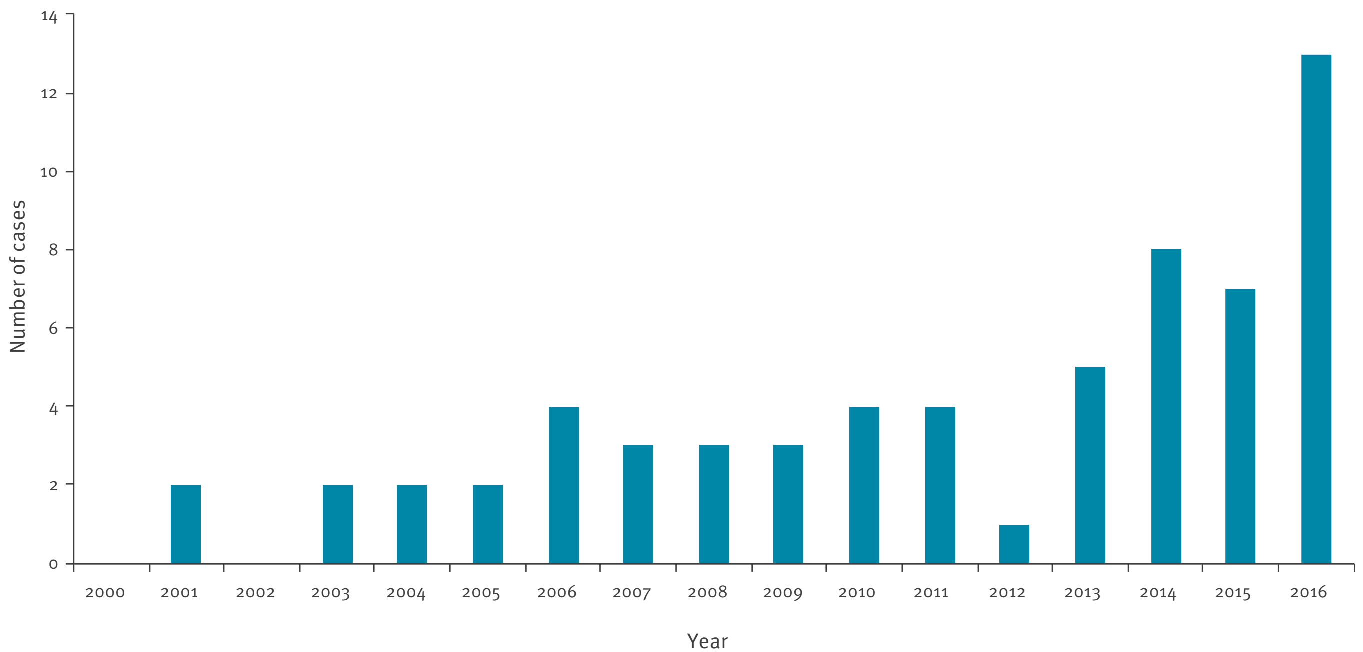
Introduction

The history of the global spread of invasive meningococcal disease (IMD) caused by serogroup W *Neisseria meningitidis* (MenW) started in the year 2000, following an international emergency during the annual Hajj season in Saudi Arabia [1]. Before that, MenW had rarely been recorded as the cause of outbreaks but rather of sporadic IMD, with a low reported incidence [1]. Recently, MenW has been spreading in different countries worldwide [2-6]. It is of concern that in the United Kingdom (UK), MenW IMD incidence has increased year by year, reaching 24% of all IMD laboratory-confirmed cases in the epidemiological year 2014/15 [5,7]. In the Netherlands, in the epidemiological year 2015/16, the MenW incidence (0.15 cases per 100,000 inhabitants) was fivefold higher than the average incidence (0.03 cases per 100,000) reported in the period from 2002/03 to 2014/15 [7].

Whole genome sequencing (WGS) evidenced the heterogeneity of meningococci belonging to serogroup W/cc11 from different geographical areas and identified several genomic types by country [5,8]. As reported by Lucidarme et al. [5,9], genomic comparison classified most of MenW/cc11 as lineage 11.1. Moreover, this lineage includes two sublineages: Hajj and South American (previously designated the 'South American/UK strain')

Figure 1

Neisseria meningitidis serogroup W causing invasive meningococcal disease, by year, Italy, 2000–2016 (n = 63)



[5,9]. The first sublineage comprises the MenW/cc11 Hajj outbreak strain, the sub-Saharan African MenW/cc11 strains from epidemic periods and the endemic South African MenW/cc11 strain [9]. The second sublineage contains three main strains: the South American strain, the original UK strain (emerged in 2009 in the UK) and the UK 2013 strain [9].

The Hajj sublineage appeared in Saudi Arabia in 2000, spreading first in the African meningitis belt and then, with smaller outbreaks, in South Africa [4,8,10]. In the UK, this sublineage caused IMD in the period from 2000 to 2004; after that, it was replaced by the endogenous MenW/cc11 strain [4,9]. In France, eight MenW/cc11 cases were reported between January and April 2012 as linked to recent travel history to Sub-Saharan Africa during the MenW epidemic [11,12].

In South America, an increase in the proportion of MenW IMD cases has also been reported in early 2000 [2]. With the exception of one IMD case reported in Brazil [3], the South American MenW/cc11 isolates were not identified as Hajj strain at that time. Later, the so-called South American sublineage was responsible for clusters in southern Brazil (2003–05), in the United States (US) (2008–09) and in Chile (2010–12) [4]. In Europe, and in particular in the UK, Ireland and France, clusters of MenW belonging to the South American strain sublineage were reported more recently, 2009–15 [8]. In Sweden, the UK 2013 strain, belonging to the South American sublineage, was the cause of an increase in MenW IMD starting from 2014 [6].

In Italy, as in the other European countries, serogroups B and C are predominant, with an increase in the proportion of isolates of serogroup Y from 2% in 2007 to 17% in 2013 [13]. Even though serogroup W has rarely been identified in the country, an increase was observed following the global spread of these meningococcal strains [7,14].

Here, the genetic variation within and between meningococci associated with invasive disease was assessed by molecular analysis of *N. meningitidis* serogroup W collected from 2000 to 2016 for an overview of the phylogenetic diversity among strains circulating in Italy. Moreover, the rapid increase in MenW cases and the contemporaneous introduction of serogroup B *N. meningitidis* (MenB) vaccine (4CMenB) into the national immunisation schedule triggered us to study the vaccine antigen genes and their genetic variability. Although this vaccine is licensed for prevention of MenB disease, the antigens are not specific to this capsular group, and a potential cross-recognition and protection against other meningococcal serogroups deserves to be evaluated.

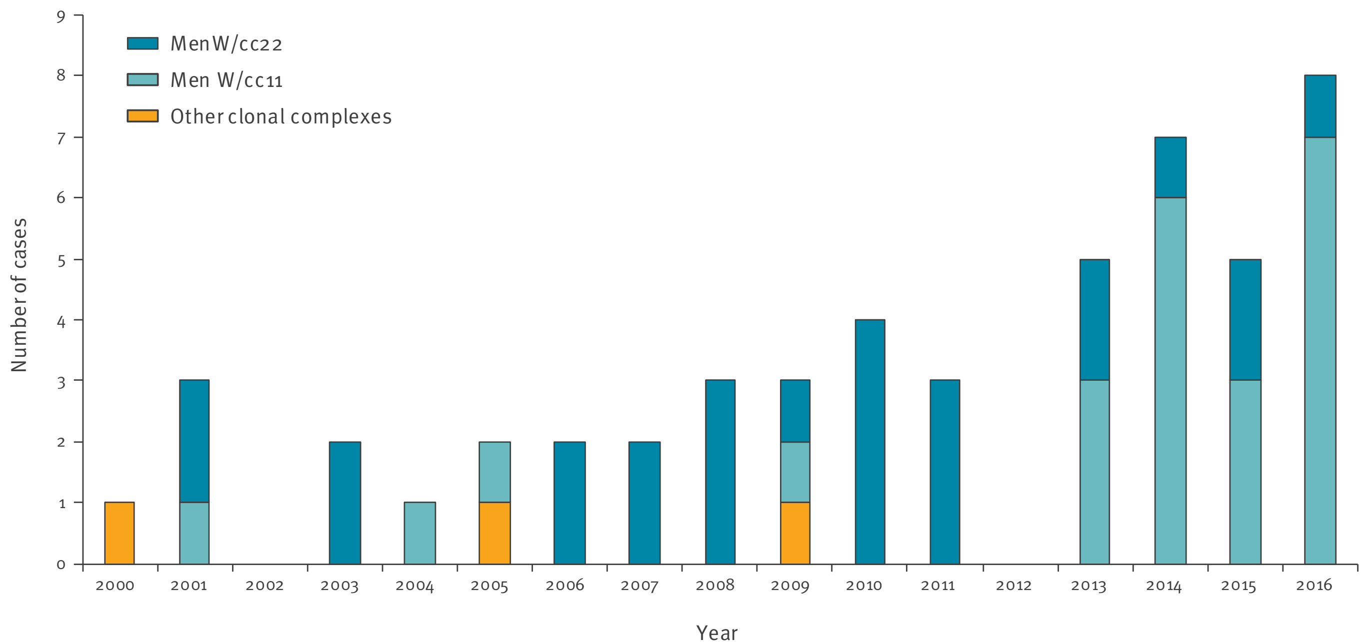
Methods

Surveillance of invasive meningococcal disease

The IMD National Surveillance System (NSS) is based on mandatory reporting to the Ministry of Health and to the Italian Institute of Public Health (Istituto Superiore di Sanità (ISS), http://www.iss.it/mabi). ISS, as national reference laboratory (NRL), acts as coordinator

Figure 2

Number of invasive meningococcal disease cases caused by *Neisseria meningitidis*, by clonal complex, Italy, 2000–2016 (n = 51)



of the NSS. Within the surveillance system, the hospital laboratories collect bacterial isolates and/or clinical samples from IMD cases and send them to the NRL for serogroup identification or confirmation and for molecular investigations. The NRL collects demographic and relevant clinical data (i.e. vaccination history) from all notified IMD cases in a dedicated database.

The data are analysed using EpiInfo software (version 3.5.3, 26 January 2011).

Microbiological analyses

For the samples sent to the NRL, the serogroup was identified or confirmed by slide agglutination with commercial antisera (Thermo Scientific, Waltham, Massachusetts, US) or by multiplex PCR [15]. For the bacterial isolates, susceptibility to cefotaxime, ceftriaxone, ciprofloxacin, penicillin G and rifampicin was determined by the minimum inhibitory concentration (MIC) test strip method (Liofilchem, Roseto degli Abruzzi, Italy) on Mueller-Hinton agar (Thermo Scientific, Waltham, Massachusetts, US) supplemented with 5% of sheep blood. The breakpoints were those recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) [16]. Chromosomal DNA was extracted using the QiAmp mini kit (Qiagen, Hilden, Germany) from an overnight culture or directly from the clinical sample, blood or cerebrospinal fluid (CSF). Multilocus sequence typing (MLST), PorA and FetA typing, MenB vaccine antigen variants and *penA* gene were identified using the PubMLST.org database (http://pubmlst.org/neisseria/). The genotypic formula comprised capsular group: *porA* (P1).VR1,VR2: *fetA* VR: ST(cc). The MenW/cc11 isolates were characterised for the allelic profile of six antigen-encoding genes (*porA, porB, fetA, nadA, nhba* and *fHbp*) suggested by Mustapha et al. as typical of the main MenW/cc11 sublineages [4].

Whole genome sequencing

Cultivated isolates were analysed by WGS. For each isolate, 1 ng of DNA was used to prepare the sequencing libraries following the Nextera XT DNA protocol. The Illumina MiSeq platform (kit v3, 600 cycles) was used for the WGS analysis. A first quality check of the raw sequence data was performed using FastQC [17]. Reads were trimmed using the software Sickle [18] to maintain a Q score > 25, and de novo assembly was carried out with the ABySS software version 1.5.2 (k parameter = 63) [19]. Contigs longer than 500 bp were selected using an ad hoc script and kept for further analysis. The final assembly ranged from 84 to 316 (median: 209) contigs per sample (N50: 10,999–59,092 bp; median: 19,790 bp), covering the ca 2.2 Mb of the *N. meningitidis* genome.

Genome comparison

Genomes were uploaded to the PubMLST.org database (http://pubmlst.org/neisseria/) and compared using the BIGSdb Genome Comparator [20] through gene-by-gene analysis. Phylogenetic analysis of the isolates was performed by core genome MLST (cgMLST) [21]. Incomplete loci were automatically removed from the distance matrix calculation for the neighbour-net graphs. The resulting distance matrices were visualised

Table

Molecular characterisation of *Neisseria meningitidis* MenW/cc11 bacterial isolates, Italy 2000–2016 (n = 18)

Bacterial isolate ID	ID (http://pubmlst.org/Neisseria)	Year of isolation	BAST	cgMLST sublineage	Six antigen-encoding gene profile
1142	42867	2001	898	Hajj	a
1505	42851	2004	898	Hajj	a
1638	42852	2005	898	Hajj	a
2205	42884	2009	898	Hajj	a
2517	42886	2013	898	Hajj	a
2693	36847	2014	898	Hajj	a
2767	42888	2015	898	Hajj	a
2808	44961	2016	898	Hajj	a
2857	51615	2016	898	Hajj	a
2904	51617	2016	898	Hajj	a
2916	51618	2016	898	Hajj	a
2940	56641	2016	898	Hajj	a
2509	42885	2013	2	South American	b
2593	36848	2014	2	South American	b
2585	36849	2014	2	South American	b
2685	36845	2015	2	South American	b
2858	51616	2016	2	South American	b
2602	42887	2013	6	Singleton	c

BAST: Bexsero antigen sequence typing; cgMLST: core genome multilocus sequence typing; ID: identification code.

as neighbour-net networks, generated by SplitsTree4 (version 4.13.1) [22].

Statistics

Change in the average annual incidence of MenW from 2000 to 2012 vs 2013 to 2016 was evaluated using a negative binomial regression model.

Results

From 2000 to 2016, 3,540 laboratory-confirmed IMD cases were reported within the NSS for IMD in Italy, with an incidence of 0.37 per 100,000 in 2016 (www.iss.it/mabi/, last access: 3 September 2018).

For 2,357 IMD cases, the capsular serogroup was identified: 1,249 were B, 861 were C, 161 were Y, 63 were W, 17 were A, five were X and one was 29E. One isolate was capsule null locus (*cnl*). The majority of cases were due to serogroups B and C, with proportions of 36% and 42%, respectively, in 2016.

As shown in Figure 1, MenW was rare from 2000 until 2012, with an average annual incidence of 0.004 per 100,000 population (30 cases). From 2013 to 2016, the average annual incidence grew to 0.01 per 100,000 population (33 cases), significantly higher than in the previous time period ($p < 0.05$). In 2016, 13 MenW cases were identified, with an incidence of 0.02 per 100,000 population, four times higher than the average value of 0.005 per 100,000 population observed in the previous years 2000 to 2015.

Among the 63 MenW IMD cases, 53 samples were sent to the NRL for further analyses: 47 bacterial isolates and six CSF or blood samples.

Demographic and clinical data of *Neisseria meningitidis* serogroup W cases

The median age of the 63 MenW cases was 20 years (mean: 29 years), ranging from 1 month to 86 years. Until 2005, MenW was responsible of IMD cases exclusively among children younger than 10 years (the median age was 1 year), except for one. In the period from 2006 to 2016, the median age increased to 26 years.

The female:male ratio was 28:35. Meningitis (25 cases) and septicaemia (22 cases) were the main clinical presentations, followed by meningitis plus septicaemia (16 cases). Four cases had an atypical clinical presentation: two (aged 3 and 26 years) had arthritis; one (20 years-old) had a pericolic abscess and one (5 months-old) had dysentery. Six patients (aged between 22 and 63 years) died, defining a case fatality rate of 10%.

Eleven patients came from foreign countries: Eritrea (n = 1) [23], Mali (n = 1) [23], Ivory Coast (n = 1), Morocco (n = 1) [23], Niger (n = 1), Nigeria (n = 5) and Somalia (n = 1).

Microbiological and molecular analyses

Antimicrobial susceptibility

Of the 47 MenW bacterial isolates received at the NRL, 44 could be cultured and tested for antimicrobial susceptibility. All of them were susceptible to cefotaxime, ceftriaxone, ciprofloxacin and rifampicin. Moreover, 14 showed decreased susceptibility to penicillin G (PenI, 0.064 › MIC ≥ 0.25) with a MIC_{50} and MIC_{90} of 0.064 mg/L and 0.19 mg/L, respectively.

MLST and genotypic formula

The molecular characterisation was performed at the NRL for 51 of 53 MenW. Two samples were not suitable for the molecular analyses. MLST identified two main clonal complexes, cc22 (n = 25) and cc11 (n = 23). In addition, two isolates were cc23 and one was cc60.

As shown in Figure 2, the main clonal complex between 2000 and 2012 was cc22 (19/26); from 2013 onward, the cc11 (19/25) was predominant. Among the 25 cc22 bacterial isolates, eight sequence types (STs) were identified: ST-22, ST-184, ST-3189, ST-904, ST-1286, ST-1959, ST-6779 and ST-11935. Among them, 12 genotypic formulas were reported, of which W:P1.18–1,3:F4–1:ST-22(cc22) was the most frequent (five bacterial isolates; Supplementary Table S1). The 23 MenW/cc11 (18 bacterial isolates and five clinical samples) belonged to ST-11 and presented a single genotypic formula, W:P1.5,2:F1–1:ST-11(cc11) (Supplementary Table S1). The cc23 isolates belonged to ST-23 and ST-9253 and the cc60 isolate to ST-913.

Whole genome sequencing

Whole genome sequencing was performed to identify: the Bexsero antigen sequence types (BAST), the cgMLST, the six antigen-encoding gene profile and the *penA* gene alleles.

BAST typing

As shown in the Table, MenW/cc11 clustered in three BAST: BAST 898 (characterised by fHbp peptide variant 1.9, NHBA peptide 96, NadA peptide 6, PorA VR1 5 and PorA VR2 2) for 12 bacterial isolates; BAST 2 (fHbp 2.22, NHBA 29, NadA 6, PorA VR1 5 and PorA VR2 2) for five; BAST 6 (fHbp 2.151, NHBA 29, NadA 6, PorA VR1 5 and PorA VR2 2) for the remaining one.

4CMenB variant antigens among MenW/cc22

The 4CMenB variant antigens identified among MenW/cc22 isolates were: fHbp peptide variant 2.16, NHBA peptide 20, NadA interrupted by an IS element. The PorA VR1,VR2 were 18–1,3 in eight isolates, 5–1,10–1 in three isolates and 5,2 in one isolate.

cgMLST

We included 1,467 of the 1,605 core genome loci in the cgMLST analysis (138 loci incompletely assembled were excluded) for 18 MenW/cc11 and seven reference genomes.

As shown in Figure 3, the 18 MenW/cc11 split into two main sublineages corresponding to those described by Lucidarme et al. [5]. Twelve genomes (ID 36847, 42851, 42852, 42867, 42884, 42886, 42888, 44961, 51615, 51617, 51618 and 56641) grouped together with reference genomes belonging to the Hajj sublineage, with a mean distance of 89 loci with allelic differences. Eight of these 12 genomes (ID 36847, 42886, 42888, 44961, 51615, 51617, 51618 and 56641) clustered in a subgroup; they had been isolated between 2013 and 2016 and five of them were associated with MenW IMD in patients with Nigerian nationality (Supplementary Table S2).

Our 12 MenW/Hajj sublineage genomes were compared with 128 MenW genomes with the genotypic formula W:P1.5,2:F1–1:ST-11(cc11) and fHbp variant 1.9, identified from IMD cases in Africa (www.neisseria.org; last accessed: 24 November 2017). All genomes showed a mean distance of 60 loci (data not shown).

Five MenW/cc11 genomes (ID 36845, 36848, 36849, 42885 and 51616) clustered together with two genomes in the South American sublineage (ID 30154, as the original UK strain reference, and ID 30167, as the UK 2013 strain reference) [9] with a mean distance of 74 loci. In particular, genomes ID 51616 and ID 30167 showed a higher proximity. The analysis of 27 of 30 genes distinguishing the original and the novel UK strains [9] confirmed that ID 51616 was a UK 2013-strain. The remaining four genomes showed a higher similarity to the original UK strain. The ID 42887 genome was close to the reference ID 21587 (South Africa 2003) in a branch far from both the main sublineages.

For the 12 MenW/cc22 genomes, 1,540 of the 1,605 core genome loci were included in the cgMLST analysis, while the remaining 65 loci were incompletely assembled. The genome comparison highlighted a mean distance of 199 loci (Supplementary Figure S1). The cgMLST analysis of MenW/cc22 and cc11 highlighted high genetic diversity with a mean distance of 588 loci (data not shown).

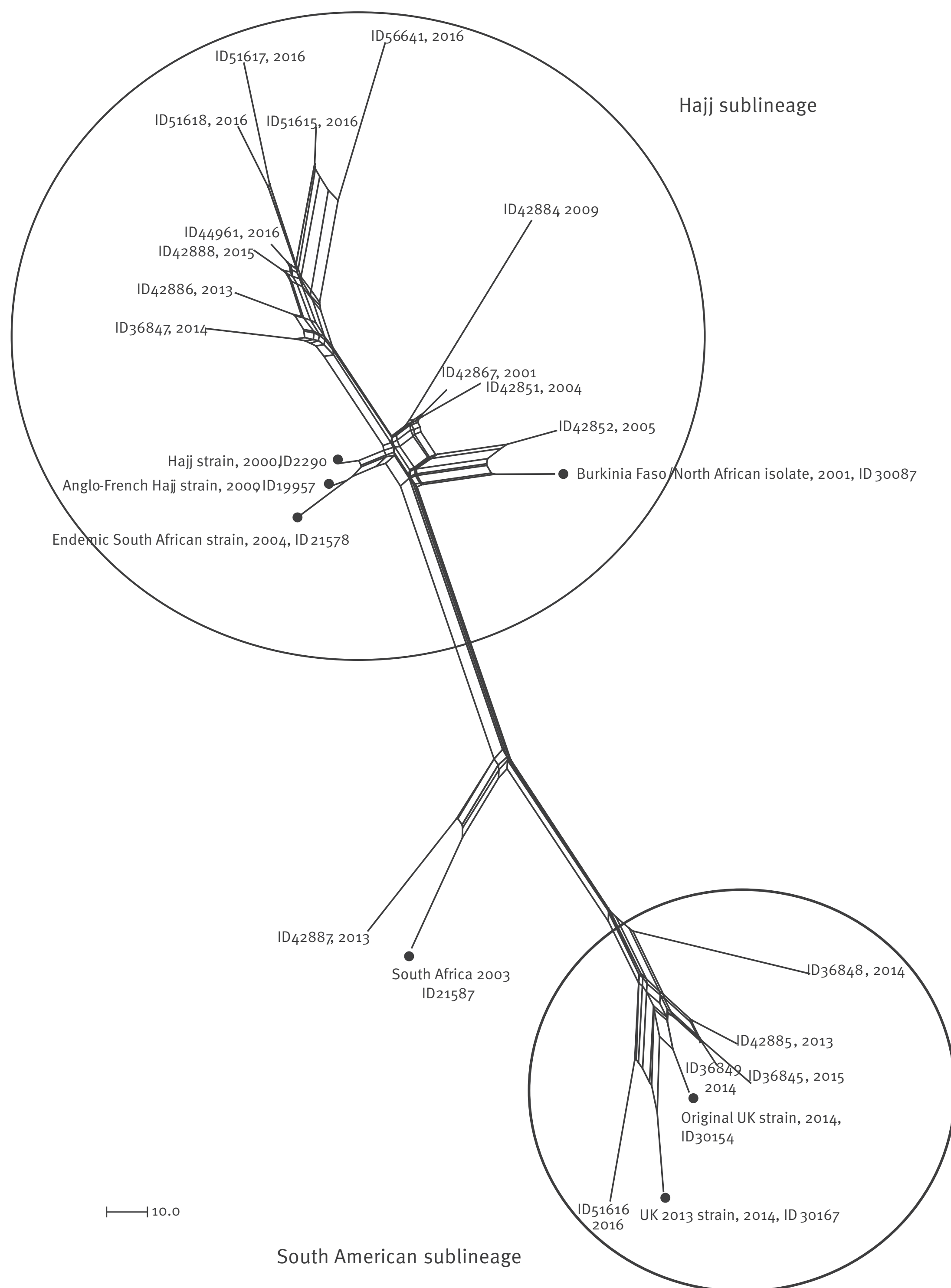
Overall, the majority of MenW/cc11 were Hajj sublineage (16/22); in particular, it caused five sporadic IMD cases from 2001 to 2013 and 11 cases from 2014 to 2016 (Supplementary Figure S2). Ten MenW/cc11 Hajj were obtained from African refugees and characterised by the presence of fHbp allele 9 (Supplementary Table S2). The South American sublineage appeared in Italy in 2013 and was responsible for five of 22 IMD cases (Supplementary Figure S2). One MenW/cc11 (ID 42887), identified in 2013, did not belong to any sublineage.

Six antigen-encoding gene profiles among MenW/cc11

Among the 18 MenW/cc11 bacterial isolates, we found three known profiles [4], comprising the alleles of *porA* , *porB*, *fetA*, *nadA*, *nhba* and *fHbp* genes (Table). Profile a was found in 12 bacterial isolates: 1, 1, 13, 5 (peptide 6), 72 (peptide 96), 9; profile b in five isolates: 1, 244,

Figure 3

Neighbour-net phylogenetic network based on a comparison of 1,467 core genome loci (cgMLST) of *Neisseria meningitidis* MenW/cc11 genomes Italy 2000–2016 (n = 25)



The figure includes genomes of 18 isolates from this study and seven reference genomes (black dots) available in the *Neisseria* PubMLST website. Last accessed: 10 April 2017. The Hajj sublineage and the South American sublineage are highlighted with circles. For each genome, the ID code from http://pubmlst.org/Neisseria is reported. The scale bar indicates the number of differences between the compared loci.

13, 5 (peptide 6), 17 (peptide 29), 22; profile c in one isolate: 1, 311, 13, 5 (peptide 6), 17 (peptide 29), 160 (peptide 151).

For four of five clinical samples, we identified *nhba* 72 and *fHbp* 9 (Supplemetary Table 2). The remaining sample was not suitable for the analysis.

As shown in the Table, the isolates clustered as Hajj sublineage showed the profile a, the isolates belonging to the South America sublineage showed the profile b and the singleton showed the profile c.

penA gene characterisation

The 18 MenW/cc11 bacterial isolates were susceptible to penicillin G and showed the *penA* allele 1. Thirteen of 24 of MenW/cc22 were PenI, of which 12 harboured the *penA* allele 14 and the remaining one the *penA* 685.

Discussion

The epidemiology of IMD is constantly changing. The national vaccination programmes should consider these changes over time and the age groups that are affected most.

Since 2000, there has been an increase in the number of MenW cases in Europe, America and Africa [2-4]. This international context prompted us to ascertain the current situation of MenW in Italy and how it had evolved over the past 17 years. Although Italy is classified as a country with a low incidence of IMD in the overall population, the number of MenW notified cases has been increasing since 2013. Data collected within the established NSS for IMD reported an increase in MenW cases, even though the absolute number was lower than that reported in other European countries [4-7]. In the past, sporadic MenW IMD cases occurred mainly among children, but have gradually increased also in older age groups, in England since the epidemiological year 2013/14, and in the Netherlands since 2015/16 [7].

In 2016, MenW represented 7% of the total IMD cases reported in Italy. In contrast to other countries [24], very few cases were characterised by atypical clinical presentation; it is likely that this is due to the small total number of reported cases and incomplete available information. In 17 years, cc11 has become the prevalent clonal complex among MenW in the country. In contrast to what was reported in Australia in 2016 [25], MenW/cc11 was not associated with the emerging resistance to penicillin.

The most interesting finding of this study is that both of the MenW/cc11 sublineages, South American and Hajj, cocirculate in Italy. Cocirculation has already been reported in some parts of the African meningitis belt and in South Africa [8], but not in Europe. In the UK in the mid-2000s, the Hajj sublineage was replaced by the South American sublineage [9]. Likewise, in France, the Hajj sublineage, detected up to 2012 [12], was replaced in 2013 by the South American sublineage [26]. The Hajj sublineage appeared in Italy in 2001 and became predominant in 2014. Across the entire study period, the Hajj sublineage represented 73% of the MenW/cc11 identified in Italy.

Five of the 22 MenW/cc11 were South American sublineage. They appeared in Italy for the first time in 2013, causing five IMD among Italian patients. Four of them were the original UK strain and only one, in 2016, was the UK 2013 strain. As extensively described, the UK 2013 strain has been spreading in northern European countries since 2013 [6,7,9].

In Italy, the National Immunisation Plan 2017–2019 recommends the quadrivalent meningococcal vaccine for adolescents, as the main group of people affected by serogroups Y and W, acting as catch-up or booster of the primary immunisation [27]. The immunisation is also recommended for travellers to countries endemic for the serogroups contained in the vaccine and for people at high risk of IMD [27]. Moreover, the recommendation for the meningococcal B vaccine (4CMenB) for infants before the age of 13 months is administered free of charge. Possible cross-protection against other non-B meningococci, through the presence of the same subcapsular vaccine antigens, need to be evaluated [28,29]. In the UK, serum bactericidal antibody (SBA) activity, promoted by immunisation with 4CMenB vaccine against *N. meningitidis* W strains, was clearly demonstrated [28]. Here, all MenW/cc11 meningococci showed the NadA peptide 6, belonging to the variant 2/3, predicted to be cross-protective with the 4CMenB NadA variant [28]. Moreover, the MenW/cc11/Hajj sublineage isolates showed the fHbp variant 1, one of the antigens of the 4CMenB vaccine. The multi-antigen typing system [30] together with SBA test [31] could define precisely the vaccine coverage against MenW; further evaluations are needed to precisely answer this question also for the MenW identified in Italy.

Conclusion

In Italy, we observed cocirculation of two sublineages, the Hajj and the South American. This is uncommon and not reported in other European countries. It is likely that the geographical location of our country may favour a peculiar epidemiological situation that needs to be carefully monitored and evaluated.

Acknowledgements

The authors thank Richard Aschbacher (Microbiology and Virology Laboratory, Azienda Sanitaria dell'Alto Adige, Bolzano, Italy), Letizia Camardese (Microbiology section, Ospedale San Carlo, Potenza, Italy), Paolo Castiglia (Università di Sassari, Sassari, Italy), Laura Daprai (Microbiology and Virology Unit, Fondazione IRCCS Ca' Grande Ospedale Maggiore, Milan, Italy), Antonino Di Caro (Microbiology section, Istituto Nazionale per le Malattie Infettive "L. Spallanzani", Rome, Italy), Teresa Fasciana (Dept. of Sciences for Health Promotion and Mother and Child Care "G. D'Alessandro", University of Palermo, Italy), Patrizia Innocenti (Microbiology and Virology Laboratory, Azienda Sanitaria dell'Alto Adige, Bolzano, Italy), Paolo Lanzafame

(Microbiology and Virology Unit, Provincial Health Services, S. Chiara Hospital, P.A. Trento, Italy), Daniela Lombardi (SeREMI, ASL AL, Alessandria, Italy), Barbara Lucignano (Microbiology Laboratory, Bambino Gesù Hospital, Rome, Italy), Maria Carla Re (Unit of Clinical Microbiology, Regional Reference Centre for Microbiological Emergencies, St. Orsola Malpighi University Hospital, Bologna, Italy) and Lucia Rossi (Microbiology and Virology Unit, University Hospital, Padova, Italy), for samples and clinical data; Maria Grazia Caporali, Fortunato Paolo D'Ancona, Martina Del Manso and Flavia Riccardo (Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy) for the collaboration in the National Surveillance System of Invasive Bacterial Diseases and for the data collection; Luigina Ambrosio and Annapina Palmieri (Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy) for technical assistance; Anna Anselmo, Antonella Fortunato, Anna Maria Palozzi, Silvia Fillo and Florigio Lista (Molecular Biology Section, Army Medical and Veterinary Research Center, Rome, Italy) for the whole genome sequencing.

This publication made use of the Neisseria Multi Locus Sequence Typing website (http://pubmlst.org/neisseria/) developed by Keith Jolley and sited at the University of Oxford. The development of this site has been funded by the Wellcome Trust and European Union.

This work was partly funded by the Italian Ministry of Health-CCM Project "Sorveglianza delle malattie invasive da *Neisseria meningitidis, Streptococcus pneumoniae e Haemophilus influenzae*" 2015 and 2016.

Conflict of interest

None declared.

Authors' contributions

Paola Stefanelli conceived the study. Cecilia Fazio provided insight on microbiological investigation and drafted the manuscript together with Paola Stefanelli. Arianna Neri and Paola Vacca carried out the laboratory analyses, contributed in the molecular analyses and provided insight into interpretation of results. Andrea Ciammaruconi carried out the whole genome sequencing. Milena Arghittu, Anna Maria Barbui, Caterina Vocale, Paola Bernaschi, Patrizia Isola, Alessandra Irene Galanti, Antonella Mencacci, Rosella De Nittis, Maria Chironna, Anna Giammanco, Elisabetta Pagani, Alessandro Bisbano were involved in the invasive meningococcal diseases surveillance at the local level. They were in charge of the data collection and management of invasive meningococcal disease cases. Paola Stefanelli revised the results. All authors participated in the drafting and revision of this manuscript and gave their final approval of this version.

References

1. Mayer LW, Reeves MW, Al-Hamdan N, Sacchi CT, Taha MK, Ajello GW, et al. Outbreak of W135 meningococcal disease in 2000: not emergence of a new W135 strain but clonal expansion within the electophoretic type-37 complex. J Infect Dis. 2002;185(11):1596-605. https://doi.org/10.1086/340414 PMID: 12023765
2. Abad R, López EL, Debbag R, Vázquez JA. Serogroup W meningococcal disease: global spread and current affect on the Southern Cone in Latin America. Epidemiol Infect. 2014;142(12):2461-70. https://doi.org/10.1017/S0950268814001149 PMID: 24831052
3. Lemos AP, Harrison LH, Lenser M, Sacchi CT. Phenotypic and molecular characterization of invasive serogroup W135 Neisseria meningitidis strains from 1990 to 2005 in Brazil. J Infect. 2010;60(3):209-17. https://doi.org/10.1016/j.jinf.2009.11.014 PMID: 20056121
4. Mustapha MM, Marsh JW, Krauland MG, Fernandez JO, de Lemos AP, Dunning Hotopp JC, et al. Genomic epidemiology of hypervirulent serogroup W, ST-11 Neisseria meningitidis. EBioMedicine. 2015;2(10):1447-55. https://doi.org/10.1016/j.ebiom.2015.09.007 PMID: 26629539
5. Lucidarme J, Hill DM, Bratcher HB, Gray SJ, du Plessis M, Tsang RSW, et al. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. J Infect. 2015;71(5):544-52. https://doi.org/10.1016/j.jinf.2015.07.007 PMID: 26226598
6. Eriksson L, Hedberg ST, Jacobsson S, Fredlund H, Mölling P, Stenmark B. Whole genome sequencing of the emerging invasive Neisseria meningitidis serogroup W in Sweden. J Clin Microbiol. 2018;56(4):e01409-17. https://doi.org/10.1128/JCM.01409-17 PMID: 29321195
7. Knol MJ, Hahné SJM, Lucidarme J, Campbell H, de Melker HE, Gray SJ, et al. Temporal associations between national outbreaks of meningococcal serogroup W and C disease in the Netherlands and England: an observational cohort study. Lancet Public Health. 2017;2(10):e473-82. https://doi.org/10.1016/S2468-2667(17)30157-3 PMID: 29253430
8. Mustapha MM, Marsh JW, Harrison LH. Global epidemiology of capsular group W meningococcal disease (1970-2015): Multifocal emergence and persistence of hypervirulent sequence type (ST)-11 clonal complex. Vaccine. 2016;34(13):1515-23. https://doi.org/10.1016/j.vaccine.2016.02.014 PMID: 26876439
9. Lucidarme J, Scott KJ, Ure R, Smith A, Lindsay D, Stenmark B, et al. An international invasive meningococcal disease outbreak due to a novel and rapidly expanding serogroup W strain, Scotland and Sweden, July to August 2015. Euro Surveill. 2016;21(45):30395. https://doi.org/10.2807/1560-7917.ES.2016.21.45.30395 PMID: 27918265
10. von Gottberg A, du Plessis M, Cohen C, Prentice E, Schrag S, de Gouveia L, et al. Emergence of endemic serogroup W135 meningococcal disease associated with a high mortality rate in South Africa. Clin Infect Dis. 2008;46(3):377-86. https://doi.org/10.1086/525260 PMID: 18181736
11. Parent du Chatelet I, Barboza P, Taha MK. W135 invasive meningococcal infections imported from Sub-Saharan Africa to France, January to April 2012. Euro Surveill. 2012;17(21):20181. PMID: 22687826
12. Taha MK, Kacou-N'douba A, Hong E, Deghmane AE, Giorgini D, Okpo SL, et al. Travel-related Neisseria meningitidis serogroup W135 infection, France. Emerg Infect Dis. 2013;19(6):1030-2. https://doi.org/10.3201/eid1906.120515 PMID: 23735310
13. Fazio C, Neri A, Renna G, Vacca P, Antonetti R, Barbui AM, et al. Persistent occurrence of serogroup Y/sequence type (ST)-23 complex invasive meningococcal disease among patients aged five to 14 years, Italy, 2007 to 2013. Euro Surveill. 2015;20(45):30061. https://doi.org/10.2807/1560-7917.ES.2015.20.45.30061 PMID: 26606870
14. Mastrantonio P, Stefanelli P, Fazio C, Sofia T, Neri A, La Rosa G, et al. Serotype distribution, antibiotic susceptibility, and genetic relatedness of Neisseria meningitidis strains recently isolated in Italy. Clin Infect Dis. 2003;36(4):422-8. https://doi.org/10.1086/346154 PMID: 12567299
15. Zhu H, Wang Q, Wen L, Xu J, Shao Z, Chen M, et al. Development of a multiplex PCR assay for detection and genogrouping of Neisseria meningitidis. J Clin Microbiol. 2012;50(1):46-51. https://doi.org/10.1128/JCM.00918-11 PMID: 22090406
16. The European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 8.1, 2018. Växjö: EUCAST; 2018. Available from: http://www.eucast.org/clinical_breakpoints/
17. Andrews S. FastQC A quality control tool for high throughput sequence data. Cambridge: Babraham Institute; 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
18. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files using quality (Version 1.33). 2011. Available from: https://github.com/najoshi/sickle
19. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19(6):1117-23. https://doi.org/10.1101/gr.089532.108 PMID: 19251739
20. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11(1):595. https://doi.org/10.1186/1471-2105-11-595 PMID: 21143983
21. Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MC. A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108

representative Neisseria meningitidis genomes. BMC Genomics. 2014;15(1):1138. https://doi.org/10.1186/1471-2164-15-1138 PMID: 25523208

22. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67. https://doi.org/10.1093/molbev/msj030 PMID: 16221896
23. Stefanelli P, Fazio C, Neri A, Rezza G, Severoni S, Vacca P, et al. Imported and Indigenous cases of Invasive Meningocococcal Disease W:P1.5,2:F1-1: ST-11 in migrants' reception centers. Italy, June-November 2014. Adv Exp Med Biol. 2016;897:81-3. https://doi.org/10.1007/5584_2015_5006 PMID: 26563305
24. Campbell H, Parikh SR, Borrow R, Kaczmarski E, Ramsay ME, Ladhani SN. Presentation with gastrointestinal symptoms and high case fatality associated with group W meningococcal disease (MenW) in teenagers, England, July 2015 to January 2016. Euro Surveill. 2016;21(12):30175. https://doi.org/10.2807/1560-7917.ES.2016.21.12.30175 PMID: 27035055
25. Mowlaboccus S, Jolley KA, Bray JE, Pang S, Lee YT, Bew JD, et al. Clonal expansion of new penicillin-resistant clade of Neisseria meningitidis serogroup W clonal complex 11, Australia. Emerg Infect Dis. 2017;23(8):1364-7. https://doi.org/10.3201/eid2308.170259 PMID: 28609259
26. Hong E, Barret AS, Terrade A, Denizon M, Antona D, Aouiti-Trabelsi M, et al. Clonal replacement and expansion among invasive meningococcal isolates of serogroup W in France. J Infect. 2018;76(2):149-58. https://doi.org/10.1016/j.jinf.2017.10.015 PMID: 29132919
27. Piano Nazionale Prevenzione Vaccinale PNPV 2017-2019 [National Plan for Vaccine Prevention NPVP 2017-2019] Rome: Ministero della Salute, 2017. Italian. Available from: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2571_allegato.pdf
28. Ladhani SN, Giuliani MM, Biolchi A, Pizza M, Beebeejaun K, Lucidarme J, et al. Effectiveness of meningococcal B vaccine against endemic hypervirulent Neisseria meningitidis W strain, England. Emerg Infect Dis. 2016;22(2):309-11. https://doi.org/10.3201/eid2202.150369 PMID: 26811872
29. Hong E, Giuliani MM, Deghmane AE, Comanducci M, Brunelli B, Dull P, et al. Could the multicomponent meningococcal serogroup B vaccine (4CMenB) control Neisseria meningitidis capsular group X outbreaks in Africa? Vaccine. 2013;31(7):1113-6. https://doi.org/10.1016/j.vaccine.2012.12.022 PMID: 23261039
30. Vogel U, Taha MK, Vazquez JA, Findlow J, Claus H, Stefanelli P, et al. Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment. Lancet Infect Dis. 2013;13(5):416-25. https://doi.org/10.1016/S1473-3099(13)70006-9 PMID: 23414709
31. Trotter CL, Findlow H, Borrow R. Seroprevalence of serum bactericidal antibodies against group W135 and Y meningococci in England in 2009. Clin Vaccine Immunol. 2012;19(2):219-22. https://doi.org/10.1128/CVI.05515-11 PMID: 22190393

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Culture-free genotyping of *Neisseria gonorrhoeae* revealed distinct strains at different anatomical sites in a quarter of patients, the Netherlands, 2012 to 2016

Brian MJW van der Veer[1], Petra FG Wolffs[1], Christian JPA Hoebe[1,2], Nicole HTM Dukers-Muijrers[1,2], Lieke B van Alphen[1]
1. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Centre (MUMC+), Maastricht, the Netherlands
2. Department of Sexual Health, Infectious Diseases and Environmental Health, South Limburg Public Health Service, Heerlen, the Netherlands

Correspondence: Lieke van Alphen (lieke.van.alphen@mumc.nl)

Citation style for this article:
van der Veer Brian MJW, Wolffs Petra FG, Hoebe Christian JPA, Dukers-Muijrers Nicole HTM, van Alphen Lieke B. Culture-free genotyping of Neisseria gonorrhoeae revealed distinct strains at different anatomical sites in a quarter of patients, the Netherlands, 2012 to 2016. Euro Surveill. 2018;23(50):pii=1800253. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800253

Article submitted on 13 May 2018 / accepted on 05 Oct 2018 / published on 13 Dec 2018

Background: Genotyping of *Neisseria gonorrhoeae* (NG) is essential for surveillance to monitor NG transmission and dissemination of resistant strains. Current genotyping methods rely on bacterial culture which frequently fails. **Aim:** Our aim was to develop a culture-free genotyping method that is compatible with the widely used *N. gonorrhoeae*multi-antigen sequence typing (NG-MAST) database, which facilitates genotyping of NG detected at separate anatomical sites in individual patients. **Methods:** Specific primers for both PCR targets *porB* and *tbpB* were designed and technically validated by assessing the analytical sensitivity, cross-reactivity with 32 non-gonoccocal Neisseria species, and concordance with NG-MAST. Clinical application was assessed on 205 paired samples from concurrent NG infections at different anatomical sites of 98 patients (81 men who have sex with men and 17 women) visiting our sexually transmitted infections clinic. **Results:** Typing could be consistently performed on samples with a PCR quantification cycle (Cq) value < 35. Furthermore, the method showed no cross-reactivity and was concordant with NG-MAST. Culture-free NG-MAST improved the typing rate from 62% (59/95) for cultured samples to 94% (89/95) compared with culture-dependent NG-MAST. Paired samples of 80 of 98 patients were genotyped, revealing distinct NG strains in separate anatomical sites in 25% (20/80) of the patients. **Conclusions:** This NG-specific genotyping method can improve NG surveillance as it facilitates genotyping of non-culturable and extra-genital samples. Furthermore, 25% of patients were infected with multiple NG strains, which is missed in current culture-dependent surveillance. Including non-culturable and concurrent NG infections in surveillance informs actions on dissemination of multidrug-resistant NG strains.

Introduction

Neisseria gonorrhoeae (NG) is one of the most common bacterial sexually transmitted infections [1]. The World Health Organization (WHO) estimates that more than 100 million new cases of NG occur each year, even though testing for NG and diagnostics have improved [1,2]. Detection of NG allows empirical treatment that results in cure in at least 95% of cases, and rapid cure subsequently limits transmission [1]. However, increasingly resistant strains of NG have been reported in the last decades, which could complicate empirical treatment [3]. Therefore, gaining insight in transmission and antimicrobial resistance (AMR) of NG is important. NG can be detected by culture or nucleic acid amplification test (NAAT) but both methods have limitations [2]. Culture is known to be less sensitive because NG requires demanding nutritional and environmental conditions, leading to a low percentage of culture-confirmed diagnoses [4]. In contrast, NAAT are more sensitive but cannot determine the AMR profile [2,5].

Surveillance of NG is essential to monitor transmission and dissemination of resistant strains. NG multi-antigen sequence typing (NG-MAST) is a widely used genotyping method to monitor transmission and outbreaks [6,7]. This method has a higher discriminatory power than multilocus sequence typing (MLST) and is more cost-effective than highly discriminatory whole genome sequencing [8,9]. In addition, some of the NG-MAST sequence types (ST) are associated with AMR [7,10]. The currently used NG-MAST protocol requires culture because the primers cross-react with other *Neisseria* species [11]. To date, only two studies have genotyped non-cultured clinical samples with NG-MAST [6,11]. Whiley et al. demonstrated that NG-MAST can be applied to non-cultured urogenital samples but not to samples from extra-genital sites

(oropharynx and rectum) because of the presence of commensal *Neisseria* species [11]. They showed that mainly *N. lactamica*, *N. meningitidis* and *N. polysaccharea* strains lead to the cross-reactivity. Furthermore, it appeared that successful application of NG-MAST to non-cultured samples was linked to the quantification cycle (Cq) of the PCR-positive sample because four of the five failed samples had a high Cq value (>35).

Previous studies have shown that patients can be NG-positive at extra-genital sites and have concurrent NG infections at different anatomical sites [12,13]. Most of the extra-genital and concurrent infections are observed in risk groups, for example in men who have sex with men (MSM). Extra-genital sites may act as a reservoir for AMR genes as the present commensal *Neisseria* species, potentially harbouring AMR genes, readily exchange DNA with NG [14]. Typing the oropharyngeal site using culture-dependent methods is especially difficult because the bacterial load is lower than at other anatomical sites and this appears to be linked to culture success [15]. In previous studies, concurrent NG infections were studied with various genotyping methods [6,11,16-18]. Distinct NG strains per anatomical site have been observed and some strains demonstrated discordant antibiotic susceptibility profiles [6,16-18]. The observed distinct NG strains could be explained by high-risk sexual behaviour and patients being part of different transmission chains [16,18]. However, these studies were small (fewer than 10 patients), focussed on cultured isolates, used a single-position single nucleotide polymorphisms (SNP) and/or used non NG-specific primers [6,16-18]. Therefore, we aimed to develop a culture-free NG-MAST genotyping method that does not cross-react with other *Neisseria* species and is compatible with the NG-MAST database. Furthermore, we aimed to gain more insight in the frequency of distinct NG strains at separate anatomical sites in individual patients.

Methods

This study was designed to test the clinical application of the culture-free NG-MAST method to non-culturable clinical samples and use these data to compare ST of separate anatomical sites within a patient. The method was technically validated by assessing analytical sensitivity, specificity and concordance with NG-MAST.

Clinical samples

All NG-positive clinical samples (n = 1,110) from different anatomical sites were retrieved from 814 consultations (further referred to as number of patients) from 642 individual patients. NG positivity was based on NG detection by the Cobas 4800 CT/NG NAAT assay (Roche Diagnostics, Basel, Switzerland), between January 2012 and May 2016 from our sexually transmitted infections (STI) clinic (South Limburg Public Health Service). These samples were from MSM (n = 769 samples), women (n = 254 samples) and heterosexual men (n = 87 samples). Samples with a Cq value of 35 or higher did not consistently yield PCR products in dilution series (see technical validation). Therefore, clinical samples with a Cq value of ≥35 were excluded (n = 418), leaving 692 samples for analysis. Of the remaining 692 NG-positive samples, we included only paired samples from separate anatomical sites belonging to a single STI clinic visit of a patient (n = 228). Different pairs of any combination of genital, anorectal or oropharyngeal NG positivity were observed. A total of 108 patients were NG-positive at two or three anatomical sites (90 MSM and 18 women who reported anal sex or symptoms and who were systematically tested on all three anatomical sites). The remaining amount of sample material was not sufficient for typing for 10 patients (nine MSM and one woman) and therefore these patients were excluded, leaving 98 patients with paired samples for analysis. In total, 205 Cobas NAAT clinical samples were included: 57 urine, 17 vaginal, 92 anorectal and 39 oropharyngeal samples. With these samples, we assessed the clinical application of the culture-free NG-MAST method and the presence of distinct STs within a patient. Data on culture success were retrieved by routine diagnostics because NG culture is mostly performed as part of the national NG resistance surveillance since NAAT diagnosis of NG is the primary diagnostic procedure. All patients were treated with a single dose of ceftriaxone, the primary choice of treatment because no resistance exists in the Netherlands [4]. An additional swab or urine sample for routine NG culture is taken at the treatment visit at the STI clinic only when treatment has not already been provided at the diagnostic visit based on symptoms. For this study, data of this routine culture was available for all patients if culture was performed.

DNA isolation clinical samples

Total DNA was isolated from 400 µL Cobas 4800 clinical samples using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) and eluted in 50 µL Milli-Q water (MQ). To increase elution yield, we extended the incubation time to 10 min. The eluate was stored at −20 °C.

DNA isolation cultured gonoccocal and non-gonoccocal *Neisseria* strains

Gonoccocal and non-gonoccocal clinical and reference *Neisseria* strains were inoculated on chocolate agar with IsoVitaleX or blood agar (BectonDickinson, Sparks, United States (US)) and incubated over night at 37 °C in 5% CO_2. Morphology of the colonies was checked and a single colony was subcultured before DNA isolation. Bacterial suspensions were prepared in sterile saline solution from two or three colonies (depending on the size of the colonies) picked with a pre-wetted sterile swab. The bacteria were pelleted by centrifugation at 2,000 g for 5 min and washed once. The pellet was resuspended in 500 µL MQ and boiled for 10 min. Cell debris was pelleted by centrifugation at 8,000 g for 2 min and the supernatant was stored at −20 °C.

Table 1

Overview of primers used in PCR and sequencing reactions for NG-MAST and culture-free NG-MAST

	NG-MAST [ ]	Culture-free NG-MAST
PCR primers *porB*		
Forward	5’-CAA GAA GAC CTC GGC AA-3’	5’-GTT AAT CCG CTA TAA CCC CC-3’
Reverse	5’-CCG ACA ACC ACT TGG T-3’	5’-CCG ACA ACC ACT TGG T-3’
PCR primers *tbpB*		
Forward	5’-CGT TGT CGG CAG CGC GAA AAC-3’	5’-TTC CTT CCA AAA AAC CGG AAG CCC G-3’
Reverse	5’-TTC ATC GGT GCG CTC GCC TTG-3’	5’-CAT TGC CCG GAT AGG CAA ACC A-3’
Sequence primers *porB*		
Forward	5’-CAA GAA GAC CTC GGC AA-3’	5’-CAA GAA GAC CTC GGC AA-3’
Reverse	5’-CCG ACA ACC ACT TGG T-3’	5’-CCG ACA ACC ACT TGG T-3’
Sequence primers *tbpB*		
Forward	5’-CGT TGT CGG CAG CGC GAA AAC-3’	5’-CGT TGT CGG CAG CGC GAA AAC-3’
Reverse	5’-TTC ATC GGT GCG CTC GCC TTG-3’	5’-TTC ATC GGT GCG CTC GCC TTG-3’

NG-MAST genotyping

PCR for both targets was performed in 50 µL reaction volumes using the Biometra T3000 Thermal Cycler (Labrepco Inc., US). Each reaction per target (*porB* and *tbpB*) contained 50 pmol of the NG-MAST forward and reverse primer for the respective target (Table 1), 2.5 U HotStar polymerase (Qiagen), 1× Qiagen PCR buffer, 0.2 mmol/L dNTP, 5 µL DNA lysate and MQ to a volume of 50 µL. The PCR protocol of Martin et al. was used to amplify the targets but cycles were increased to 30 [7].

The amplicons were precipitated with 50 µL 20% polyethylene glycol 8000 and 2.5 mol/L sodium chloride at 37 °C for 15 min. Precipitated amplicons were centrifuged at 15,000 × g for 15 min and washed twice with ice-cold 80% ethanol. The pellet was allowed to dry and resuspended in 25 µL MQ.

The *porB* and *tbpB* fragments were sequenced with their respective forward and reverse primer using the BigDye Terminator v1.1 Cycle Sequencing kit (Thermo Fisher Scientific, Waltham, Massachusetts, US). The sequence protocol has an initial denaturation step of 1 min at 96 °C, followed by 25 cycles of 10 s at 96 °C, 10 s at 55 °C (*porB*) or 65 °C (*tbpB*), and 3 min at 60 °C.

Primer design culture-free NG-MAST

The genome sequences of all NG reference strains published by the WHO (n = 14) were downloaded from GenBank and used for multiple alignments with Clustal Omega [19]. A 2 kb flanking region of the aligned NG-MAST *porB* and *tbpB* primers were selected to identify conserved regions. Each flanking region was aligned and conserved regions were tested for in silico specificity using basic-local alignment search tool (BLAST). A specific sequence was identified that could be used as the forward primer for *porB* but no specific sequence was identified for the reverse primer, therefore the NG-MAST reverse primer was used which resulted in a fragment of ca 1.2 kb (Table 1). Two specific sequences were identified for *tbpB* which could be used as a forward and reverse primer, resulting in a fragment of ca 1.8 kb (Table 1).

Culture-free NG-MAST genotyping

This method was similar to the NG-MAST method apart from the initial PCR. Each reaction per target (*porB* and *tbpB*) contained 50 pmol of the culture-free NG-MAST forward and reverse primer for the respective target (Table 1), 0.2 µL AccuPrime Taq DNA Polymerase High Fidelity (Thermo Fisher), 1× AccuPrime PCR buffer II, 15 µL DNA isolated from a clinical sample and MQ to a volume of 50 µL. The PCR protocol had an initial denaturation step of 5 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 60 s at 58 °C (*porB)* or 69 °C (*tbpB*), 2.5 min at 68 °C, and a final extension of 10 min at 68 °C. The *porB* and *tbpB* amplicons were sequenced with NG-MAST primers (Table 1). The culture-free method was therefore compatible with the NG-MAST online database because we characterised the same fragments of *porB*and *tbpB* genes.

Technical validation of culture-free NG-MAST method

Analytical sensitivity was determined using dilution series ranging from 1.3×10^6 to 1.3×10^2 colony-forming units (CFU)/mL. Concordance of culture-free NG-MAST method with NG-MAST was tested with seven randomly selected isolates cultured from four urine samples, two anorectal swabs and one oropharyngeal swab, and their respective unculturable Cobas 4800 screening samples

between January and March 2017. The isolates were subjected to NG-MAST genotyping whereas the clinical samples were subjected to the culture-free NG-MAST method. The analytical specificity was tested with a panel of 32 non-gonococcal *Neisseria* species strains, including *N. cinerea* (n = 1), *N. denitrificans* (n = 1), *N. elongata* (n = 1), *N. flavescens* (n = 1), *N. lactamica* (n = 2), *N. meningitidis* (n = 3), *N. mucosa* (n = 7), *N. perflava* (n = 1), *N. polysaccharea* (n = 1) and *N. subflava* (n = 14).

Data analysis

The trace files were assembled, trimmed and edited using Bionumerics (version 7.6, Applied Maths, Sint-Martens-Latem, Belgium). The starting trimming patterns for *porB* and *tbpB* and lengths were used as described in Martin et al. [7]. Alleles and ST were called according to the NG-MAST online curated database. Phylogenetic trees of *porB*, *tbpB*and concatenated sequences were constructed using multiple alignment and unweighted pair group method with arithmic mean (UPGMA) clustering using default settings with gap penalty at 100%.

Ethical statement

The study protocol was approved as a scientific study not done in humans by the Medical Ethical Committee of Maastricht University Medical Centre (MUMC+; number METC 2017–2-0251) as it concerned a laboratory and observational study using anonymous data and leftover diagnostic samples only. This was part of an STI clinic procedure where patients did not object to the use of their data and samples anonymously for research purposes.

Results

Analytical sensitivity, specificity and concordance of culture-free NG-MAST

Dilution series in triplicate showed that culture-free NG-MAST consistently yielded PCR products for both *porB* and *tbpB* in samples with a Cq value <35. None of the tested 32 non-gonococcal *Neisseria* strains were PCR-positive for either *porB* or *tbpB* in the culture-free NG-MAST PCR reactions. The seven randomly selected cultured isolates had identical ST as their respective unculturable Cobas 4800 screening sample but distinct ST were observed between the selected isolates.

Paired clinical samples

In total, 90.2% (185/205) of the selected paired clinical samples were successfully genotyped with the culture-free NG-MAST method. The *porB* fragment was successfully sequenced in 95.6% (196/205) of samples and *tbpB* in 93.7% (192/205). Failure of both targets in a sample does not appear to be related to the Cq value because both low (<30) and higher (30–35) Cq values show comparable failure rates (data not shown). We observed 36 different *porB* and 22 *tbpB* alleles, resulting in 45 ST. Among the samples, *porB*-1808 and *tbpB*-29 were the most common alleles, present in 51 and 49 samples, respectively. Furthermore, we found five previously unidentified *porB* and two *tbpB* alleles which all had the highest identity with NG using a BLAST search. The most prevalent ST were ST2992 (n = 36), ST11461 (n = 30), and ST5441 (n = 26), and 15 new STs were found.

Routine culture was performed for 95 of the 205 paired clinical samples and 59 (62.1%) were culture-positive. Typically, only one anatomical site was sampled for culture, and the majority of the culture-positive samples were collected from the genital site (44/59). Culture-free NG-MAST applied to the non-culturable clinical material (Cobas 4800 sample material) of samples send in for culture (including culture-negative samples) showed that 93.7% (89/95) were genotyped successfully. However, four samples negative in culture-free NG-MAST were culture-positive. Of the remaining 110 uncultured clinical samples, 98 (89.1%) could be genotyped.

Sequence diversity within *por*B and *tbp*B alleles

High sequence diversity was observed for both *porB* and *tbpB* in this study population (Supplement Figures S1 and S2). Two *porB* alleles (90 and 2723) were divergent, with more than 50% dissimilarity, from all other observed alleles. The newly identified *tbpB* allele with 91% similarity with *tbpB*-1251 was divergent from all other observed alleles with more than 60% dissimilarity. In addition, the average dissimilarity between *tbpB* alleles appeared to be greater than between *porB* alleles.

Sequence types of samples from separate anatomical sites in a patient

In this dataset of clinical samples, we genotyped 169 paired samples (taken from a single patient at separate anatomical sites) from 80 patients (66 MSM and 14 women) (Supplement Table). We observed distinct concurrent ST in a quarter (20/80) of the patients. They had the following combinations of sample material: urine-anorectal (n = 6), urine-oropharyngeal (n = 1), anorectal-oropharyngeal (n = 8), urine-anorectal-oropharyngeal (n = 1), vaginal-anorectal (n = 3), and vaginal-anorectal-oropharyngeal (n = 1) (Table 2). Similar proportions of distinct concurrent ST were observed in MSM (16/66) and women (4/14). Interestingly, a single patient (patient 32) was NG-positive with a distinct NG strain at all three tested anatomical sites (Table 2).

The Figure presents the dissimilarity of concatenated sequences of *porB* and *tbpB* between STs. For the majority of the patients with distinct concurrent STs, a large (>15%) dissimilarity was observed between the concatenated sequences. Patients 31 and 48 had only 1% dissimilarity between the concatenated sequences. In both patients, the *tbpB* allele was identical between the distinct ST but the *porB* allele showed >1% dissimilarity, meaning the two ST did not belong to the same genogroup. When assigning ST to genogroups,

Table 2

Characteristics of patients with concurrent *Neisseria gonorrhoeae* infection with distinct sequence types, including age, risk group, multiple sequence types and NG-MAST results per sample site, the Netherlands, January 2012–May 2016 (n = 42)

Patient	Age (years)	Risk group	Urine			Vaginal			Anorectal			Oropharyngeal		
			porB	tbpB	ST	porB	tbpB	ST	porB	tbpB	ST	porB	tbpB	ST
9	26	Women[a]	-	-	-	1808	91%-1251	New ST1	1808	29	2992	-	-	-
10	22	MSM	-	-	-	-	-	-	3031	33	4995	1489	33	10257
24	27	MSM	-	-	-	-	-	-	7272	33	New ST4	99%-7988	110	New ST3
27	24	Women[a]	-	-	-	301	29	359	2723	110	4431	301	29	359
28	20	MSM	-	-	-	-	-	-	908	27	3588	2723	27	15046
29	41	MSM	6720	188	11461	-	-	-	1808	137	11084	-	-	-
31	44	MSM	7988	110	13902	-	-	-	908	110	1407	-	-	-
32	42	MSM	182	74	1247	-	-	-	1808	29	2992	1808	836	New ST6
33	21	Women[a]	-	-	-	99%-6405	74	New ST7	6720	188	11461	-	-	-
37	25	MSM	1808	188	New ST8	-	-	-	182	74	1247	-	-	-
39	20	MSM	-	-	-	-	-	-	3031	33	4995	4288	4	9382
48	21	MSM	-	-	-	-	-	-	6720	188	11461	4288	188	New ST9
53	41	MSM	30	18	5441	-	-	-	3059	29	5049	-	-	-
54	25	MSM	-	-	-	-	-	-	1808	29	2992	30	18	5441
55	36	Women[a]	-	-	-	1808	29	2992	6720	188	11461	-	-	-
58	49	MSM	30	18	5441	-	-	-	-	-	-	1808	29	2992
59	26	MSM	-	-	-	-	-	-	1808	29	2992	30	18	5441
60	42	MSM	30	18	5441	-	-	-	1808	29	2992	-	-	-
79	20	MSM	-	-	-	-	-	-	4199	29	New ST14	4199	4	New ST12
80	19	MSM	6720	188	11461	-	-	-	30	29	298	-	-	-

NG-MAST: *Neisseria gonorrhoeae* multi-antigen sequence typing; ST: sequence type.

[a] Women reporting anal sex or symptoms.

Sample material that was NG-negative or not sampled as part of routine diagnostics is indicated with a hyphen.

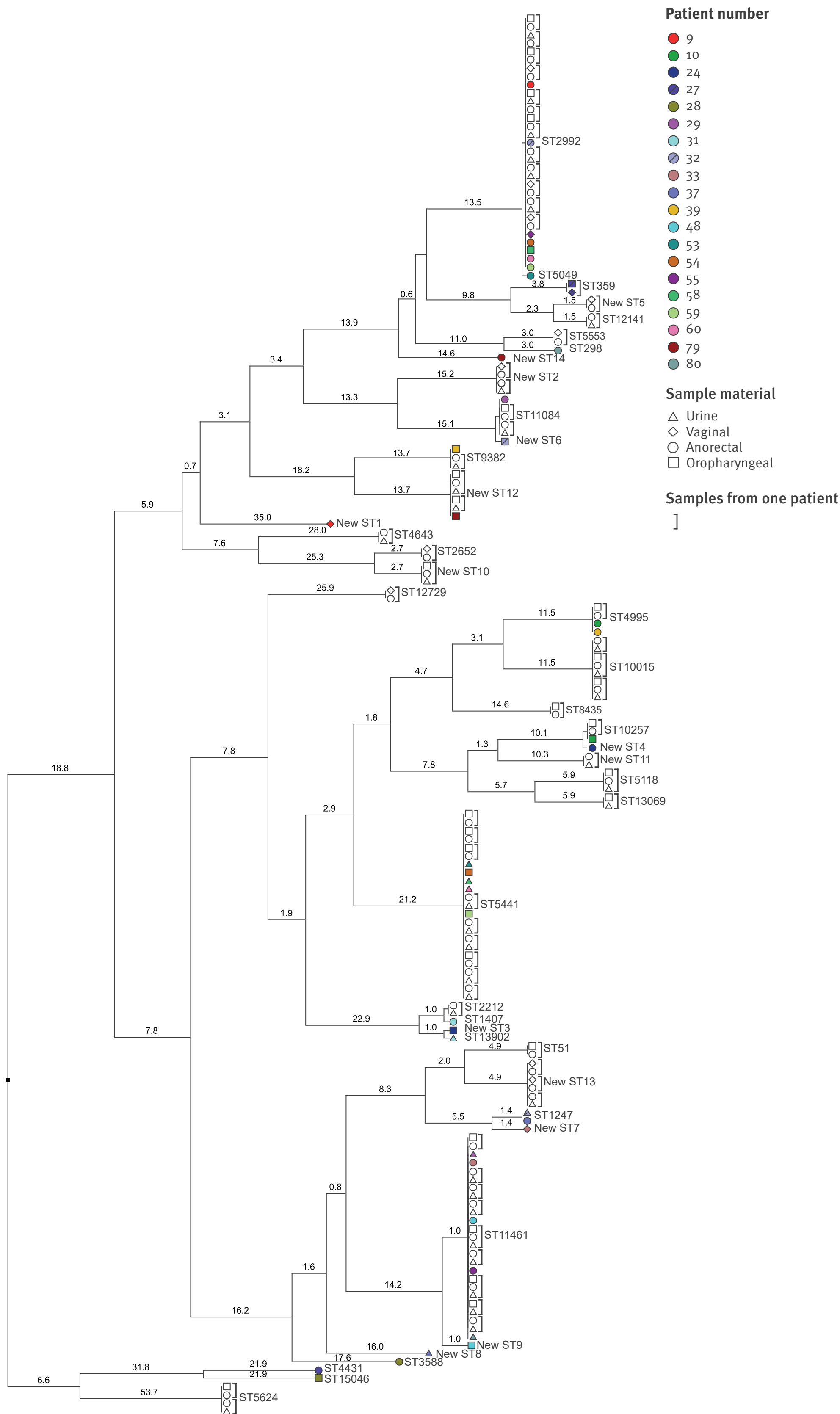
we identified two genogroups that consisted of more than five samples: G2992 (n=32) and G11084 (n=6). Furthermore, we identified three samples belonging to G1407 (one ST1407 and two ST2212) of which only one (ST2212) was culture-positive and susceptible for ceftriaxone.

Discussion

In this study we show that the culture-free NG-MAST method can readily be used to genotype NG in clinical samples including extra-genital samples. In addition, the method is compatible with the online NG-MAST database. The culture-free NG-MAST method was technically validated by assessing the NG analytical specificity using non-gonoccocal *Neisseria* species; it demonstrated good specificity for NG as no cross-reactivity was observed. Furthermore, concordance with NG-MAST was demonstrated by comparing typing results of non-culturable clinical samples (Cobas 4800 sample material) with cultured isolates, which were taken less than 2 weeks apart. In this time frame, identical genotypes were expected based on the study by Martin et al. [7]. With culture-free NG-MAST, we genotyped 90% of the selected paired clinical samples with sufficient bacterial load (Cq value <35). Extrapolating this genotyping rate to all NG positive samples (n=1,110) would result in successful typing of 56% (624/1,110) of all NG-positive samples. Among all samples sent in for culture, culture-free NG-MAST showed a higher typing rate of 94% (89/95) compared with the culture-dependent method with 62% (59/95). However, four of the 59 culture-positive samples were negative in culture-free NG-MAST. The clinical samples testing negative in culture-free NG-MAST could be caused by PCR-inhibitory substances in the clinical material. The majority of the culture-positive samples were collected from the genital site (44/59); that could be explained by the sampling strategy, but the low sensitivity of extra-genital NG culture could also have contributed [2]. This highlights the importance of culture-free genotyping as the current surveillance data would be biased towards genital samples. With

Figure

Dendogram constructed by multiple alignment of concatenated *porB* and *tbpB* sequences clustered with unweighted pair group method with arithmic mean (UPGMA) algorithm, the Netherlands, January 2012–May 2016 (n = 169)



Sequence type (ST) of each sample is shown at the tips of the dendogram and samples of patients with distinct ST are coloured. The branch length depicts the percentage of dissimilarity.

our method, we were able to genotype 33 of the 36 culture-negative samples, which were mainly extra-genital samples (25/33).

Our results show that both sexes were frequently infected with distinct NG strains in a quarter of patients (20/80) which is higher than most previous studies [6,11,17,18]. The studies of Whiley et al. (0/4) and Carannante et al. (1/8) assessed, respectively, only four and eight patients with paired samples, which could explain the lower proportion [6,11]. Pond et al. (3/71) developed a real-time PCR assay to predict ciprofloxacin resistance with the detection of a resistance-associated SNP [17]. This method uses a single position to identify distinct strains of NG, leading to a lower resolution than NG-MAST where two internal fragments (490 bp and 390 bp) of highly polymorphic genes are analysed. De Silva et al. (26/206) performed whole genome sequencing only on cultured strains and therefore may have missed distinct strains from samples that were NAAT-positive but culture-negative [18]. A higher percentage of distinct strains in paired clinical samples was reported in a study by Kolader et al. (52/130) which applied *por-opa*restriction fragment length polymorphism typing [16]. The authors hypothesised that the observed high frequency could be the result of high-risk sexual behaviour or also of recombination in the *opa* genes.

High-risk behaviour and sex with multiple sex partners on the same occasion may explain the frequently observed distinct ST in our study as we included MSM and women reporting anal sex or symptoms attending our STI clinic, who are considered as risk groups [1,16]. Another possible reason could be DNA exchange with commensal *Neisseria*species or other NG strains [11,16]. Patients colonised with multiple NG strains could have different AMR profiles, potentially resulting in under-treatment which could allow dissemination of resistant strains [6,17]. However, the impact of multiple strain infections on treatment needs to be addressed in future research to answer questions of the effect on AMR development and dissemination of resistant strains.

The observed concurrent infections with distinct strains in our study would be overlooked in routine diagnostics as Dutch and European (European Centre for Disease Prevention and Control) NG resistance surveillance guidelines recommend culture of only one anatomical site [2,20,21]. Without typing data for concurrent NG infections, surveillance data are incomplete and potential transmission links or associations between ST and AMR can be missed. This potentially results in dissemination of unrecognised resistant types. Therefore, early detection and improved surveillance of ST that are linked to AMR could minimise sequelae and prevent dissemination of multidrug-resistant strains.

We observed high variability in both alleles and ST in our study population, which could be due to sampling over a prolonged time period and from different risk groups (MSM and women reporting anal sex or symptoms). For example, ST2212, ST2992, ST5441 and ST5793 are more prevalent in MSM than heterosexual men or women [10,22,23]. In our study, these STs were mainly found in MSM, but eight samples with ST2992 and one with ST5793 were from women. Interestingly, we found three samples which belong to the genogroup G1407 (ST1407 (n = 1) and ST2212 (n = 2)) linked to decreased susceptibility to the last first-line treatment with ceftriaxone [10,23]. Only one of the three could be cultured (ST2212) and was susceptible to ceftriaxone. Furthermore, ST359, ST2992, ST3588 and ST4995 are linked to azithromycin resistance which is the recommended dual-therapy treatment with ceftriaxone in case of a *Chlamydia trachomatis* co-infection [1,23-25]. In addition, this dual therapy is applied to slow down emerging resistance or where local resistance data are not available [1,5,20]. In the Netherlands, a single treatment with ceftriaxone is applied because no resistance has yet been found in the Netherlands [4]. However, a multidrug-resistant isolate was recently found in the United Kingdom that showed high-level resistance to both ceftriaxone and azithromycin, thereby highlighting the need for improved surveillance [26]. In our study population, we found a high prevalence of ST belonging to genogroup G2992 (19.5%), which is in line with earlier data from the Netherlands (16.1%), while genogroups G1407 (1.7%) and G359 (1.1%) were less frequent (respectively 7.7% and 6.3% in the study of Wind et al.) [23]. The genogroup G2992 is also frequently observed in most other countries in Europe [27] and G1407 prevalence is higher in most European countries than in the Netherlands.

Even though many NG-MAST ST are linked to resistance profiles in NG, this does not necessarily imply that the strain is phenotypically resistant [7,10]. Additional tests that can identify mutations leading to resistance, for example azithromycin resistance, could give more insight into those strains that cannot be cultured [28]. A limitation of this study is that we only included samples with a higher NG load (Cq value <35); therefore the typing rate of samples with a lower bacterial load is unknown. However, as culture success is also associated with bacterial load, culture-dependent methods are expected to perform worse than our culture-free method in samples with a low NG load (Cq value ≥35). This hypothesis is strengthened because only 10% (18/188) of the samples with a Cq value ≥35 were culture-positive in routine diagnostics. A nested PCR approach might improve genotyping of samples with a low bacterial load as has been applied for medico legal purposes to allow typing from a piece of clothing [29].

Conclusion

The culture-free NG-MAST method can genotype NG from most non-culturable clinical samples, including extra-genital samples, as cross-reactivity with commensal *Neisseria* species was not observed. Compared with culture-dependent NG-MAST, culture-free

NG-MAST has a higher typing rate and does not have high demands on sample conditions. Applying culture-free NG-MAST to clinical samples revealed frequent concurrent infections with distinct ST at separate anatomical sites in MSM and women reporting anal sex or symptoms. These distinct concurrent ST and extragenital NG infections would be missed in the current European surveillance strategy possibly allowing dissemination of resistant NG strains. Including non-culturable and concurrent NG infections in surveillance informs actions to contain the dissemination of multidrug-resistant NG strains.

Conflict of interest

None declared.

Authors' contributions

Conceptualisation: PW, CH, LvA. Formal analysis: BvdV, PW, LvA. Methodology: BvdV, PW, LvA. Supervision: PW, CH, NDM, LvA. Validation: BvdV, PW. Writing - Original draft: BvdV. Writing - Review and editing: PW, CH, NDM, LvA.

References

1. World Health Organization (WHO). Global action plan to control the spread and impact of antimicrobial resistance in Neisseria gonorrhoeae. Geneva: WHO; 2012. Available from: http://apps.who.int/iris/bitstream/handle/10665/44863/9789241503501_eng.pdf;jsessionid=DFA2F277BB3490186A525ABFEF62B3E5?sequence=1
2. Recommendations for the laboratory-based detection of Chlamydia trachomatis and Neisseria gonorrhoeae--2014. MMWR Recomm Rep. 2014;63(RR-02):1-19.
3. Goire N, Lahra MM, Chen M, Donovan B, Fairley CK, Guy R, et al. Molecular approaches to enhance surveillance of gonococcal antimicrobial resistance. Nat Rev Microbiol. 2014;12(3):223-9. https://doi.org/10.1038/nrmicro3217 PMID: 24509781
4. Visser M, van Aar F, Op de Coul ELM, Slurink IAL, van Wees DA, Hoenderboom BM, et al. Sexually transmitted infections in the Netherlands in 2017. Bilthoven: National Institute for Public Health and Environment; 2018. Available from: https://www.rivm.nl/bibliotheek/rapporten/2018-0012.pdf
5. World Health Organization (WHO). WHO guidelines for the treatment of Neisseria gonorrhoeae. Geneva: WHO; 2016. Available from: http://apps.who.int/iris/bitstream/handle/10665/246114/9789241549691-eng.pdf?sequence=1
6. Carannante A, Ghisetti V, Dal Conte I, Gregori G, Stella ML, Vacca P, et al. Molecular characterization of Neisseria gonorrhoeae on non-cultured specimens from multiple anatomic sites. Ann Ist Super Sanita. 2017;53(3):213-7. PMID: 28956800
7. 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. https://doi.org/10.1086/383047 PMID: 15073688
8. Unemo M, Dillon JA. Review and international recommendation of methods for typing neisseria gonorrhoeae isolates and their implications for improved knowledge of gonococcal epidemiology, treatment, and biology. Clin Microbiol Rev. 2011;24(3):447-58. https://doi.org/10.1128/CMR.00040-10 PMID: 21734242
9. Kwong JC, Gonçalves da Silva A, Dyet K, Williamson DA, Stinear TP, Howden BP, et al. NGMASTER:in silico multi-antigen sequence typing for Neisseria gonorrhoeae. Microb Genom. 2016;2(8):e000076. PMID: 28348871
10. 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):20358. PMID: 23351652
11. Whiley DM, Goire N, Ray ES, Limnios A, Lambert SB, Nissen MD, et al. Neisseria gonorrhoeae multi-antigen sequence typing using non-cultured clinical specimens. Sex Transm Infect. 2010;86(1):51-5. https://doi.org/10.1136/sti.2009.037689 PMID: 19843535
12. Dukers-Muijrers NH, Schachter J, van Liere GA, Wolffs PF, Hoebe CJ. What is needed to guide testing for anorectal and pharyngeal Chlamydia trachomatis and Neisseria gonorrhoeae in women and men? Evidence and opinion. BMC Infect Dis. 2015;15(1):533. https://doi.org/10.1186/s12879-015-1280-6 PMID: 26576538
13. Chan PA, Janvier M, Alexander NE, Kojic EM, Chapin K. Recommendations for the diagnosis of Neisseria gonorrhoeae and Chlamydia trachomatis, including extra-genital sites. Med Health R I. 2012;95(8):252-4. PMID: 22970467
14. Unemo M, Shafer WM. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. Clin Microbiol Rev. 2014;27(3):587-613. https://doi.org/10.1128/CMR.00010-14 PMID: 24982323
15. Bissessor M, Tabrizi SN, Fairley CK, Danielewski J, Whitton B, Bird S, et al. Differing Neisseria gonorrhoeae bacterial loads in the pharynx and rectum in men who have sex with men: implications for gonococcal detection, transmission, and control. J Clin Microbiol. 2011;49(12):4304-6. https://doi.org/10.1128/JCM.05341-11 PMID: 21956992
16. Kolader ME, Dukers NH, van der Bij AK, Dierdorp M, Fennema JS, Coutinho RA, et al. Molecular epidemiology of Neisseria gonorrhoeae in Amsterdam, The Netherlands, shows distinct heterosexual and homosexual networks. J Clin Microbiol. 2006;44(8):2689-97. https://doi.org/10.1128/JCM.02311-05 PMID: 16891479
17. Pond MJ, Hall CL, Miari VF, Cole M, Laing KG, Jagatia H, et al. Accurate detection of Neisseria gonorrhoeae ciprofloxacin susceptibility directly from genital and extragenital clinical samples: towards genotype-guided antimicrobial therapy. J Antimicrob Chemother. 2016;71(4):897-902. https://doi.org/10.1093/jac/dkv432 PMID: 26817487
18. De Silva D, Peters J, Cole K, Cole MJ, Cresswell F, Dean G, et al. Whole-genome sequencing to determine transmission of Neisseria gonorrhoeae: an observational study. Lancet Infect Dis. 2016;16(11):1295-303. https://doi.org/10.1016/S1473-3099(16)30157-8 PMID: 27427203
19. Unemo M, Golparian D, Sánchez-Busó L, Grad Y, Jacobsson S, Ohnishi M, et al. The novel 2016 WHO Neisseria gonorrhoeae reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. J Antimicrob Chemother. 2016;71(11):3096-108. https://doi.org/10.1093/jac/dkw288 PMID: 27432602
20. Workowski KA, Bolan GACenters for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm Rep. 2015;64(RR-03):1-137. PMID: 26042815
21. European Centre for Disease Prevention and Control (ECDC). Gonococcal antimicrobial susceptibility surveillance in Europe 2015. Stockholm: ECDC; 2017. Available from: https://ecdc.europa.eu/sites/portal/files/documents/gonococcal-antimicrobial-susceptibility-surveillance-Europe-2015.pdf
22. Cheng CW, Li LH, Su CY, Li SY, Yen MY. Changes in the six most common sequence types of Neisseria gonorrhoeae, including ST4378, identified by surveillance of antimicrobial resistance in northern Taiwan from 2006 to 2013. J Microbiol Immunol Infect. 2016;49(5):708-16. https://doi.org/10.1016/j.jmii.2014.08.016 PMID: 25442864
23. Wind CM, Bruisten SM, Schim van der Loeff MF, Dierdorp M, de Vries HJC, van Dam AP. A Case-Control Study of Molecular Epidemiology in Relation to Azithromycin Resistance in Neisseria gonorrhoeae Isolates Collected in Amsterdam, the Netherlands, between 2008 and 2015. Antimicrob Agents Chemother. 2017;61(6):e02374-16. https://doi.org/10.1128/AAC.02374-16 PMID: 28373191
24. Shigemura K, Osawa K, Miura M, Tanaka K, Arakawa S, Shirakawa T, et al. Azithromycin resistance and its mechanism in Neisseria gonorrhoeae strains in Hyogo, Japan. Antimicrob Agents Chemother. 2015;59(5):2695-9. https://doi.org/10.1128/AAC.04320-14 PMID: 25712352
25. Brunner A, Nemes-Nikodem E, Jeney C, Szabo D, Marschalko M, Karpati S, et al. Emerging azithromycin-resistance among the Neisseria gonorrhoeae strains isolated in Hungary. Ann Clin Microbiol Antimicrob. 2016;15(1):53. https://doi.org/10.1186/s12941-016-0166-9 PMID: 27646968
26. Eyre DW, Sanderson ND, Lord E, Regisford-Reimmer N, Chau K, Barker L, et al. Gonorrhoea treatment failure caused by a Neisseria gonorrhoeae strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. Euro Surveill. 2018;23(27):1800323. https://doi.org/10.2807/1560-7917.ES.2018.23.27.1800323 PMID: 29991383

27. European Centre for Disease Prevention and Control (ECDC). Molecular typing of Neisseria gonorrhoeae - a study of 2013 isolates. Stockholm: ECDC; 2018. Available from: https://ecdc.europa.eu/sites/portal/files/documents/Molecular-typing-N-gonorrhoeae-web.pdf
28. Trembizki E, Buckley C, Donovan B, Chen M, Guy R, Kaldor J, et al. Direct real-time PCR-based detection of Neisseria gonorrhoeae 23S rRNA mutations associated with azithromycin resistance. J Antimicrob Chemother. 2015;70(12):3244-9. PMID: 26338048
29. Martin IM, Foreman E, Hall V, Nesbitt A, Forster G, Ison CA. Non-cultural detection and molecular genotyping of Neisseria gonorrhoeae from a piece of clothing. J Med Microbiol. 2007;56(Pt 4):487-90. https://doi.org/10.1099/jmm.0.46956-0 PMID: 17374888

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Research article

Improvement of Legionnaires’ disease diagnosis using real-time PCR assay: a retrospective analysis, Italy, 2010 to 2015

Maria Luisa Ricci[1], Antonella Grottola[2,5], Giulia Fregni Serpini[2], Antonino Bella[1], Maria Cristina Rota[1], Francesca Frascaro[2], Emanuela Pegoraro[3], Marisa Meacci[2], Anna Fabio[2], Elena Vecchi[4], Antonietta Girolamo[1], Fabio Rumpianesi[2], Monica Pecorari[2], Maria Scaturro[1]

1. Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy
2. Unit of Microbiology and Virology, Polyclinic University Hospital, Modena, Italy
3. U.O.C. of Microbiology and Virology, Azienda Ospedaliero-Universitaria, Verona, Italy
4. Hospital Hygiene, Polyclinic University Hospital, Modena, Italy
5. Department of Surgery, Medicine, Dentistry and Morphological Scientists with Transplant Surgery, Oncology and Regenerative Medicine Relevance, University of Modena and Reggio Emilia, Modena, Italy

Correspondence: Maria Scaturro (maria.scaturro@iss.it)

Citation style for this article:
Ricci Maria Luisa, Grottola Antonella, Fregni Serpini Giulia , Bella Antonino, Rota Maria Cristina, Frascaro Francesca, Pegoraro Emanuela, Meacci Marisa, Fabio Anna, Vecchi Elena, Girolamo Antonietta, Rumpianesi Fabio, Pecorari Monica, Scaturro Maria. Improvement of Legionnaires’ disease diagnosis using real-time PCR assay: a retrospective analysis, Italy, 2010 to 2015. Euro Surveill. 2018;23(50):pii=1800032. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800032

Article submitted on 16 Jan 2018 / accepted on 06 Aug 2018 / published on 13 Dec 2018

Aim: To evaluate real-time PCR as a diagnostic method for Legionnaires’ disease (LD). Detection of *Legionella* DNA is among the laboratory criteria of a probable LD case, according to the European Centre for Disease Prevention and Control, although the utility and advantages, as compared to culture, are widely recognised. **Methods:** Two independent laboratories, one using an in-house and the other a commercial real-time PCR assay, analysed 354 respiratory samples from 311 patients hospitalised with pneumonia between 2010–15. The real-time PCR reliability was compared with that of culture and urinary antigen tests (UAT). Concordance, specificity, sensitivity and positive and negative predictive values (PPV and NPV, respectively) were calculated. **Results:** Overall PCR detected eight additional LD cases, six of which were due to *Legionella pneumophila* (Lp) non-serogroup 1. The two real-time PCR assays were concordant in 99.4% of the samples. Considering in-house real-time PCR as the reference method, specificity of culture and UAT was 100% and 97.9% (95% CI: 96.2–99.6), while the sensitivity was 63.6% (95%CI: 58.6–68.6) and 77.8% (95% CI: 72.9–82.7). PPV and NPV for culture were 100% and 93.7% (95% CI: 91.2-96.3). PPV and NPV for UAT were 87.5% (95% CI: 83.6-91.4) and 95.8% (95% CI: 93.5-98.2). **Conclusion:** Regardless of the real-time PCR assay used, it was possible to diagnose LD cases with higher sensitivity than using culture or UAT. These data encourage the adoption of PCR as routine laboratory testing to diagnose LD and such methods should be eligible to define a confirmed LD case.

Introduction

Legionnaires’ disease (LD) is a severe form of pneumonia and is caused by bacteria belonging to the *Legionella* genus. These microorganisms are ubiquitous in natural freshwater environments and can also thrive in man-made water systems. *Legionella pneumophila* (Lp) is the mostly responsible for the development of LD; serogroup 1 (sg1) is most frequently isolated from clinical samples [1]. LD cannot be clinically or radiologically distinguished from pneumonia cases of different aetiology, therefore the disease often remains undiagnosed. Age, underlying diseases, delay in diagnosis and inappropriate antibiotic therapy can result in an increase of the case fatality rate from LD [2].

In 2015, the enumeration of all cases with a known outcome demonstrated an average case fatality rate of 8%, with a higher rate (28%) in nosocomial cases in Europe [3]. According to LD case definition [4,5], culture, a fourfold raise in Lp sg1 antibodies and urinary antigen test (UAT) are the only laboratory methods considered reliable for LD case confirmation. While serology has been nearly abandoned, UAT has almost completely replaced culture, representing 82% and 97% of diagnosis in Europe and in the United States (US), respectively [1,3]. A similar trend was observed in Italy, where in 2016 UAT and culture were used to diagnose 95.5% and 2.7% of cases, respectively [6]. However, both culture and UAT have some limitations; culture is time consuming and has a sensitivity ranging from < 10–80% [1], UAT can be performed rapidly and with very high specificity for Lp sg1, but sensitivity for non-sg1 antigens is very low. In addition, the sensitivity

Table 1
Clinical samples analysed for admitted patients, Italy, 2010–15 (n = 311)

	Number of tested samples	Number of positive samples	Number of negative samples	Number of individuals tested
In-house real-time PCR assay	354	55	299	311
Commercial real-time PCR assay	354	53	301	311
UAT	278	40	238	246
Culture	354	35	319	311

PCR: polymerase chain reaction; UAT: urinary antigen test.

of UAT has been demonstrated to be lower for non-Lp sg1 MAb 3/1-positive strains [2]. Of note, laboratory diagnosis is often based on a single method, without taking into account the limitations that each diagnostic assay might have [3,7].

Diagnostic tools based on detection of nucleic acids are an option to overcome the limitations observed by both culture and UAT. The numerous PCR assays proposed have shown high sensitivity and specificity, provided fast results and were able to detect a higher number of cases, giving the possibility to improve surveillance and better characterise local LD epidemiology [8-14]. Despite an increase in the proportion of cases diagnosed by PCR being reported in several European countries, the use of PCR is still very limited; presently a positive PCR result only defines a LD probable case [4,5]. Currently, in Italy, only 0.1% of LD cases are diagnosed by PCR [6].

The aim of this retrospective study was to evaluate real-time PCR as rapid diagnostic tool to define a LD case.

Methods

Respiratory samples were analysed using two different real-time PCR assays, performed in two different laboratories.

Samples collection

A total of 369 respiratory samples (including sputa, bronchial-alveolar lavages and bronchial aspirates) collected from 326 patients admitted to hospital for any pneumonia between 2010 and 2015 in Italy and were retrospectively analysed for *Legionella pneumophila* DNA detection.

Clinical samples were collected by two hospital laboratories, 74 samples (from 74 patients) from the Laboratory of Microbiology and Virology (University Hospital of Verona) and 295 (from 252 patients) from the Modena Regional Reference Laboratory (RRL) for Clinical Diagnosis of Legionellosis (Unit of Microbiology and Virology-Polyclinic University Hospital). All clinical samples were obtained 1 or 2 days after the onset of symptoms except three samples that were collected 5 days after onset of the disease. After collection, respiratory samples were stored at - 80 °C until tested.

Furthermore, 278 urine samples were available from 246 patients. There were 74 urine samples from 74 patients from Verona and 204 urine samples from 172 patients from Modena RRL.

Culture examination and urinary antigen test

While patients were hospitalised with pneumonia symptoms, the Laboratory of Microbiology and Virology of Verona and the Modena RRL performed *Legionella* culture and UAT. For 25 patients culture was performed on two different respiratory samples and for nine patients on three samples, while for the remaining patients culture was performed on only one sample. Culture was carried out according to the procedures described elsewhere [15].

Both laboratories performed UAT by using both BinaxNOW Legionella Urinary Antigen Card kit and Binax Legionella Urinary Antigen EIA kit (Alere, Scarborough, US). Urine samples were always boiled before testing. For 19 patients UAT was performed on two urinary samples and for eight patients on three samples.

Real-time PCR

DNA extraction was performed at the Modena RRL using the ELITe STAR 200 Extraction kit (ELITechGroup S.p.A, Torino, Italy). DNA extracts were split in two aliquots to be analysed by real-time PCR at the Modena RRL and at the National Reference Laboratory at the Istituto Superiore di Sanità in Rome.

The Modena RRL analysed 5µL of DNA with the CE IVD marked real-time PCR commercial kit Legionella pn. Q-PCR Alert (ELITechGroup, CE IVD marked) detecting for Lp *mip* gene, according to the manufacturer's instructions on a 7300 Real-Time PCR System (Applied Biosystems, Foster City, California (CA), US). The NRL also analysed 5µL of DNA using an in-house real-time PCR assay in a final volume of 20µL, containing 10µL of EXPRESS qPCR SuperMix, (Invitrogen, Carlsbad, CA, US), with Chromo 4 BioRad instrument (Bio-Rad, Hercules, CA, US), updated to CFX-96, and the following thermal protocol: 5 minutes at 95°C followed by

Table 2

Legionnaires' disease cases with at least one positive diagnostic test, Italy, 2010–15 (n = 52)

Culture	Urinary antigen test	In-house real-time PCR	Number of cases N = 52
Positive	Positive	Positive	21
Positive	Negative	Positive	3
Positive	ND	Positive	5
Negative	Positive	Positive	10
Negative	ND	Positive	3
Negative	Negative	Positive	5
Negative	Positive	Positive	5

ND: not done.

45 cycles of denaturation at 95°C for 10 seconds and annealing/extension at 60 °C for 15 seconds. Primers and probes were as described by Mentasti et al. [14], targeting *mip* and *wzm* genes for detection of Lp (sg1–15) and sg1 marker, respectively. Primers and probes for internal control DNA were also as already described [14].

Statistical analysis

The concordance between tests was evaluated using the Kappa test (K < 0.20 = "poor"; 0.20–0.40 = "fair"; 0.40–0.60 = "moderate"; 0.60–0.80 = "good"; 0.80–1.00 = "very good"). The specificity, sensitivity and the positive and negative predictive values (PPV and NPV, respectively) and 95% confidence intervals (CI) for all methods were calculated considering the in-house real-time PCR as a reference method. In addition, the concordance between all methods was also calculated. All statistical analyses were performed by Stata software version 11.2 (StataCorp, Texas, US).

Results

Samples analysed by real-time PCR, culture and urinary antigen test

Of 369 DNA samples, 15 were excluded from the comparison with culture and UAT because they were inhibitory in both PCR assays, as demonstrated by the absence of amplification of the internal control. These samples were also found negative for culture and UAT. Therefore, 354 samples from 311 patients were included in the comparison of PCR results with culture and/or UAT results (Table 1).

Both commercial and in-house real-time PCR assays gave the same results in 352 out of 354 samples, of which 299 (85%) were negative and 53 (15%) were positive (53 positive for *mip* marker and six positive also for *wzm* target). The in-house PCR detected two more positive samples (n = 55) compared with the commercial one. Of the 354 samples analysed by in-house PCR, six samples, (five negatives for both culture and UAT and one negative only for UAT but positive for culture) were identified as Lp non-sg1. Since the in-house PCR assay was able to differentiate Lp sg1 from the other serogroups, it was arbitrarily considered as a reference assay.

The concordance of the two PCR assays (commercial vs in-house) was 99.4% with a K = 0.98 (p < 0.0001). Specificity and sensitivity of commercial PCR assay were calculated equal to 100% and 96.4% (95% CI: 94.4–98.3) respectively.

All 354 respiratory samples were also tested by culture; of these, 35 (9.9%) were positive.

A total of 278 urine samples were tested by UAT and 40 (14.3%) were found positive. The two methods used to detect the urinary antigen were concordant on all tested samples.

Legionnaires' disease cases detected

The total number of LD cases detected was 52 (Table 2) and it was calculated considering the patients with at least one positive diagnostic test (culture, UAT and PCR). The in-house PCR assay was considered as a reference for comparison with culture and UAT results.

Using culture and/or urinary antigen test for diagnosis, the number of LD cases detected was 44; when the in-house PCR assay was added, the number of detected cases increased to 52 (Table 2). PCR confirmed LD diagnosis in 84.6% of cases with at least one traditional diagnostic criterion positive (culture or UAT or both) and an increment of 18.2% was observed.

The comparison between culture and the in-house real-time PCR assay showed that the sensitivity of culture (63.6%; 95% CI: 58.6–68.6) was lower, while the specificity was 100%. The PPV and the NPV were 100% and 93.7% (95% CI: 91.2-96.3), respectively. Overall concordance was good (94.3%; k = 0.75; p < 0.0001) (Table 3).

The comparison between UAT and the in-house PCR showed a higher sensitivity (77.8%; 95% CI: 72.9–82.7) than between culture and PCR, while specificity was slightly lower (97.9%; 95% CI: 96.2–99.6), and PPV and NPV were 87.5% (95% CI: 83.6–91.4) and 95.8% (95% CI: 93.5–98.2) respectively. Overall concordance of the two assays was good (94.6%; k = 0.79; p < 0.0001) (Table 3).

Discussion

In this study two independent laboratories, using a different real-time PCR assay for *Legionella pneumophila* DNA detection, analysed 354 respiratory samples and provided results with a very high concordance (99.4%).

Our results highlight a higher sensitivity of PCR compared with culture and a higher diagnostic efficiency compared with UAT. Furthermore, as recently stressed

Table 3

Comparison of culture and UAT vs in-house real-time PCR by sensitivity, specificity, PPV, NPV, concordance and kappa value, Italy, 2010–15

Comparison	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Concordance (%)	Kappa value	p-value
Culture vs real-time PCR in-house	63.6 (58.6–68.6)	100.0	100.0	93.7 (91.2–96.3)	94.3	0.75	< 0.0001
UAT vs real-time PCR in-house	77.8 (72.9–82.7)	97.9 (96.2–99.6)	87.5 (83.6–91.4)	95.8 (93.5–98.2)	94.6	0.79	< 0.0001

CI: confidence interval; NPV: negative predictive value; PCR: polymerase chain reaction; PPV: positive predictive value; UAT: urinary antigen test.

by other authors [16,17], it is important to perform more than one diagnostic assay in order to properly diagnose LD. Five of the eight LD cases with negative UAT results would have been missed if PCR assays, able to detect all Lp serogroups, had not been performed. Although in some instances UAT can incidentally detect non-1 Lp serogroups, they are designed to specifically detect Lp1 antigen, therefore, negative UAT results do not completely rule out LD infection. In addition to the aforementioned five cases (negative for UAT and for culture), three more culture-negative cases, resulted positive for Lp DNA by PCR. For these three, clinicians had only requested cultures and did not request UAT. Overall the eight additional cases show that even with a negative diagnosis but in presence of pneumonia, LD infection should be suspected and all available tests performed to investigate it.

Considering that urine samples were boiled before testing to destroy heat-sensitive proteins that could affect the test, false positive results can be reasonably excluded [7]. A possible explanation for the five UAT-positive but PCR-negative cases was obtained querying patients' records: for two patients a sputum sample was promptly collected and analysed, while for the others sputum analysis was requested 5 or more days after the antibiotic therapy was started. Although there are not sufficient data to show if and how PCR results might be affected by an on-going antibiotic therapy, the above observation suggests the need to perform PCR assay as soon as possible, ideally before or immediately after the initiation of the antibiotic treatment.

The NPV was suggestive of the excellent reliability of the PCR methods, even though only Lp DNA was targeted. However, this limitation can often be found also using culture method, because specific and selective *Legionella*isolation media, such as buffered charcoal yeast extract (BCYE) and glycine vancomycin polymyxin cycloheximide (GVPC), poorly support *Legionella* non-*pneumophila* growth [18]. The PPV was also consistent with a higher sensitivity of PCR than culture.

The reliability of PCR in diagnosing LD is more and more recognised by the scientific community and recent studies demonstrated a better performance of PCR compared with other diagnostic assays, regardless of the type of respiratory sample (bronchoalveolar lavage or sputum) [9,13]. Moreover, PCR can also detect the presence of all *Legionella* species some of which are notoriously difficult to isolate by culture [19].

In this study, the use of real-time PCR resulted in an increment of eight (18.2%) identified LD cases and therefore is an objective improvement in the diagnosis of LD. Real-time PCR has been considered a poorly reliable method due to the risk of cross-contaminations, however, the introduction of automated procedures for DNA extractions and also for PCR set up, has resulted in a consistent improvement in preventing this PCR drawback. Therefore, after an appropriate validation of their own molecular tests, clinical microbiology laboratories can adopt PCR assays to detect *Legionella* in respiratory samples.

The adoption of rapid methods to quickly identify LD cases is a priority, as the infection rate is underestimated all over the world and difficult to quantify, and increasing in several countries [2,3,20]. The laboratory procedures currently used to define confirmed LD cases are not able to guarantee a high level of sensitivity and specificity of results and they can be time-consuming. As a rapid LD diagnosis is crucial for both patient management and public health purposes, real-time PCR should be considered and implemented both locally and at *Legionella* reference laboratories in combination with all the other available methods.

In conclusion, as already observed in other countries, this study shows that the introduction of real-time PCR can improve LD diagnosis and should be considered among the criteria to define confirmed cases of LD [13].

Acknowledgements

We are grateful to Massimo Mentasti for his precious comments and editing of the manuscript.

This work was supported by the Ministry of Health (Centro per il controllo delle malattie, 2013-2014; Grant N. 5M12).

Conflict of interest

None declared.

Authors' contributions

Maria Scaturro, Maria Luisa Ricci, Monica Pecorari and Antonella Grottola designed the study. Antonella Grottola, Giulia Fregni Serpini, Francesca Frascaro, Antonietta Girolamo and Maria Scaturro performed the real-time PCR experiments. Antonella Grottola, Giulia Fregni Serpini, Francesca Frascaro, Monica Pecorari, Maria Luisa Ricci and Maria Scaturro elaborated the results. Emanuela Pegoraro provided respiratory samples and analysed them by culture; Emanuela Pegoraro also provided urinary antigen test data. Marisa Meacci and Anna Fabio analysed respiratory samples by culture and urine by urinary antigen test. Elena Vecchi and Monica Pecorari queried patients' records. Antonino Bella and Maria Cristina Rota performed the statistical analysis. Maria Scaturro wrote the manuscript and Antonella Grottola and Giulia Fregni Serpini helped her with the editing. All authors read and approved the final version of the manuscript.

References

1. Mercante JW, Winchell JM. Current and emerging Legionella diagnostics for laboratory and outbreak investigations. Clin Microbiol Rev. 2015;28(1):95-133. https://doi.org/10.1128/CMR.00029-14 PMID: 25567224
2. Phin N, Parry-Ford F, Harrison T, Stagg HR, Zhang N, Kumar K, et al. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis. 2014;14(10):1011-21. https://doi.org/10.1016/S1473-3099(14)70713-3 PMID: 24970283
3. European Centre for Disease Prevention and Control (ECDC). Annual Epidemiological report for 2015. Legionnaires' disease. Stockholm: ECDC; 2017. Available from: https://ecdc.europa.eu/en/publications-data/legionnaires-disease-annual-epidemiological-report-2015
4. European Centre for Disease Prevention and Control (ECDC). European Union (EU) case definitions. Stockholm: ECDC; 2018. Available from: https://ecdc.europa.eu/en/infectious-diseases-public-health/surveillance-and-disease-data/eu-case-definitions
5. Centers for Disease Control and Prevention (CDC). 2012 National Notifiable Infectious Diseases (Historical). Atlanta: CDC; 2012. Available from: https://wwwn.cdc.gov/nndss/conditions/notifiable/2012/infectious-diseases/
6. L'Istituto Superiore di Sanità (ISS). La legionellosi in Italia nel 2017. [Legionellosis in Italy in 2017]. Rome: ISS; 2018. Italian. Available from: http://old.iss.it/binary/publ/cont/ONLINE.pdf
7. Rota MC, Fontana S, Montaño-Remacha C, Scaturro M, Caporali MG, Vullo V, et al. Legionnaires' disease pseudoepidemic due to falsely positive urine antigen test results. J Clin Microbiol. 2014;52(6):2279-80. https://doi.org/10.1128/JCM.00493-14 PMID: 24719437
8. Avni T, Bieber A, Green H, Steinmetz T, Leibovici L, Paul M. Diagnostic Accuracy of PCR Alone and Compared to Urinary Antigen Testing for Detection of Legionella spp.: a Systematic Review. J Clin Microbiol. 2016;54(2):401-11. https://doi.org/10.1128/JCM.02675-15 PMID: 26659202
9. Botelho-Nevers E, Grattard F, Viallon A, Allegra S, Jarraud S, Verhoeven P, et al. Prospective evaluation of RT-PCR on sputum versus culture, urinary antigens and serology for Legionnaire's disease diagnosis. J Infect. 2016;73(2):123-8. https://doi.org/10.1016/j.jinf.2016.04.039 PMID: 27306488
10. Chen DJ, Procop GW, Vogel S, Yen-Lieberman B, Richter SS. Utility of PCR, Culture, and Antigen Detection Methods for Diagnosis of Legionellosis. J Clin Microbiol. 2015;53(11):3474-7. https://doi.org/10.1128/JCM.01808-15 PMID: 26292304
11. Benitez AJ, Winchell JM. Clinical application of a multiplex real-time PCR assay for simultaneous detection of Legionella species, Legionella pneumophila, and Legionella pneumophila serogroup 1. J Clin Microbiol. 2013;51(1):348-51. https://doi.org/10.1128/JCM.02510-12 PMID: 23135949
12. Murdoch DR. Diagnosis of Legionella infection. Clin Infect Dis. 2003;36(1):64-9. https://doi.org/10.1086/345529 PMID: 12491204
13. Murdoch DR, Podmore RG, Anderson TP, Barratt K, Maze MJ, French KE, et al. Impact of routine systematic polymerase chain reaction testing on case finding for Legionnaires' disease: a pre-post comparison study. Clin Infect Dis. 2013;57(9):1275-81. https://doi.org/10.1093/cid/cit504 PMID: 23899682
14. Mentasti M, Kese D, Echahidi F, Uldum SA, Afshar B, David S, et al. Design and validation of a qPCR assay for accurate detection and initial serogrouping of Legionella pneumophila in clinical specimens by the ESCMID Study Group for Legionella Infections (ESGLI). Eur J Clin Microbiol Infect Dis. 2015;34(7):1387-93. https://doi.org/10.1007/s10096-015-2363-4 PMID: 25851812
15. Istituto Superiore di Sanità (ISS). Linee guida per la prevenzione ed il controllo della legionellosi [Guidelines for prevention and control of legionellosis]. Rome: ISS; 2015. Italian. Available from: http://old.iss.it/binary/iss4/cont/C_17_pubblicazioni_2362.pdf
16. Peci A, Winter AL, Gubbay JB. Evaluation and Comparison of Multiple Test Methods, Including Real-time PCR, for Legionella Detection in Clinical Specimens. Front Public Health. 2016;4:175. https://doi.org/10.3389/fpubh.2016.00175 PMID: 27630979
17. Svarrer CW, Lück C, Elverdal PL, Uldum SA. Immunochromatic kits Xpect Legionella and BinaxNOW Legionella for detection of Legionella pneumophila urinary antigen have low sensitivities for the diagnosis of Legionnaires' disease. J Med Microbiol. 2012;61(2):213-7. https://doi.org/10.1099/jmm.0.035014-0 PMID: 21921112
18. Lee JV, Lai S, Exner M, Lenz J, Gaia V, Casati S, et al. An international trial of quantitative PCR for monitoring Legionella in artificial water systems. J Appl Microbiol. 2011;110(4):1032-44. https://doi.org/10.1111/j.1365-2672.2011.04957.x PMID: 21276147
19. Vaccaro L, Izquierdo F, Magnet A, Hurtado C, Salinas MB, Gomes TS, et al. Correction: First Case of Legionnaire's Disease Caused by Legionella anisa in Spain and the Limitations on the Diagnosis of Legionella non-pneumophila Infections. PLoS One. 2016;11(9):e0162934. https://doi.org/10.1371/journal.pone.0162934 PMID: 27607064
20. Farnham A, Alleyne L, Cimini D, Balter S. Legionnaires' disease incidence and risk factors, New York, New York, USA, 2002-2011. Emerg Infect Dis. 2014;20(11):1795-802. https://doi.org/10.3201/eid2011.131872 PMID: 25513657

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Research article

Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples

Liana E. Kafetzopoulou[1,2], Kyriakos Efthymiadis[3], Kuiama Lewandowski[1], Ant Crook[1], Dan Carter[1,2], Jane Osborne[4], Emma Aarons[4], Roger Hewson[1,2], Julian A. Hiscox[2,5], Miles W. Carroll[1,2], Richard Vipond[1,2], Steven T. Pullan[1,2]
1. Public Health England, National Infections Service, Porton Down, United Kingdom
2. NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, United Kingdom
3. Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium
4. Rare and Imported Pathogens Laboratory, Public Health England, Porton Down, United Kingdom
5. Institute of Infection and Global Health, University of Liverpool, United Kingdom

Correspondence: Steven Pullan (steven.pullan@phe.gov.uk)

Citation style for this article:
Kafetzopoulou Liana E., Efthymiadis Kyriakos, Lewandowski Kuiama, Crook Ant, Carter Dan, Osborne Jane, Aarons Emma, Hewson Roger, Hiscox Julian A., Carroll Miles W., Vipond Richard, Pullan Steven T.. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Euro Surveill. 2018;23(50):pii=1800228. https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800228

Article submitted on 27 Apr 2018 / accepted on 23 Oct 2018 / published on 13 Dec 2018

Background: The recent global emergence and re-emergence of arboviruses has caused significant human disease. Common vectors, symptoms and geographical distribution make differential diagnosis both important and challenging. **Aim:** To investigate the feasibility of metagenomic sequencing for recovering whole genome sequences of chikungunya and dengue viruses from clinical samples. **Methods:** We performed metagenomic sequencing using both the Illumina MiSeq and the portable Oxford Nanopore MinION on clinical samples which were real-time reverse transcription-PCR (qRT-PCR) positive for chikungunya (CHIKV) or dengue virus (DENV), two of the most important arboviruses. A total of 26 samples with a range of representative clinical Ct values were included in the study. **Results:** Direct metagenomic sequencing of nucleic acid extracts from serum or plasma without viral enrichment allowed for virus identification, subtype determination and elucidated complete or near-complete genomes adequate for phylogenetic analysis. One PCR-positive CHIKV sample was also found to be coinfected with DENV. **Conclusions:** This work demonstrates that metagenomic whole genome sequencing is feasible for the majority of CHIKV and DENV PCR-positive patient serum or plasma samples. Additionally, it explores the use of Nanopore metagenomic sequencing for DENV and CHIKV, which can likely be applied to other RNA viruses, highlighting the applicability of this approach to front-line public health and potential portable applications using the MinION.

Introduction

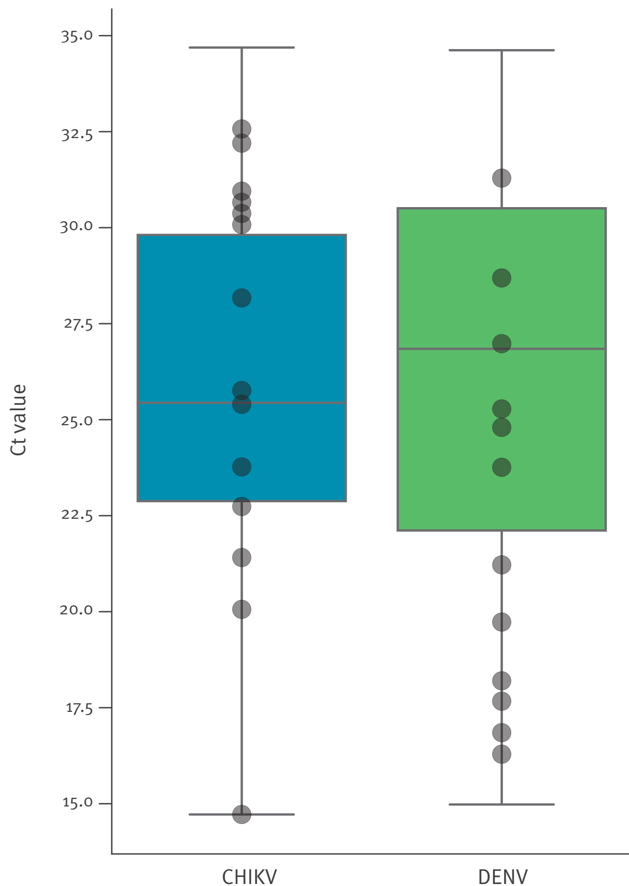
Arboviruses are predominantly RNA viruses that replicate in haematophagous (blood-sucking) arthropod vectors such as ticks, mosquitoes and other biting flies to maintain their transmission cycle [1]. Human disease outbreaks caused by arboviruses have increased in prevalence since the 2000's, led by the spread of mosquito-borne arboviruses such as chikungunya (CHIKV), dengue (DENV), West Nile (WNV), yellow fever (YFV) and Zika (ZIKV) viruses across both hemispheres [2]. CHIKV and DENV are of particular global health concern, as they have lost the need for enzootic amplification and consequently have caused extensive epidemics [3].

CHIKV is a single-stranded positive-sense RNA virus of the alphavirus genus, which causes the debilitating arthritic disease, chikungunya [4]. It has spread globally and been designated a serious emerging disease by the World Health Organization [5]. Outbreaks of CHIKV since 2005 have been associated with increased morbidity and possibly mortality [6,7].

DENV, which causes dengue, is a single-stranded positive-sense RNA virus of the flavivirus genus and the most prevalent human arboviral pathogen. Dengue occurs following infection with one of four DENV serotypes (DENV1–4). A minority of cases develop acute haemorrhagic manifestations and multi-organ failure. Despite DENV cases being under-reported, a 143.1% increased global incidence was estimated between 2005 and 2015 [8]. Approximately 500,000 DENV infected patients worldwide require hospitalisation annually [9].

Figure 1

Cycle threshold (Ct) values distribution of chikungunya (n = 73) and dengue virus (n = 368) positive samples from the Rare and Imported Pathogens Laboratory, Public Health England, United Kingdom, 2016 (n = 441 total samples)



CHIKV: chikungunya virus; Ct: cycle threshold; DENV: dengue virus.

The 14 CHIKV and 12 DENV samples selected for this work are indicated by circles. For each virus, the median Ct value of positive samples by quantitative real-time PCR is shown (horizontal line inside the box), as well as 25th and 75th percentiles (box lower and upper boundaries) and total range (whiskers).

Both CHIKV and DENV are predominantly transmitted to humans via *Aedes* species mosquitoes, particularly *Ae. aegypti* and *Ae. albopictus* [10,11], and share clinical presentations of arthralgia, headache, high fever, myalgia and rash. Circulation of CHIKV, DENV (and other arboviruses) in the same areas leads to challenges in differential diagnosis, especially in endemic regions in which diagnosis is predominantly symptom-based [12]. Additionally, reports of arboviral coinfections are increasingly common [13-16].

Metagenomic RNA sequencing allows for identification of multiple pathogens within a sample in a non-targeted and unbiased manner. It has identified causative agents in outbreaks, e.g. Lujo virus in South Africa [17], Bundibugyo ebolavirus in Uganda [18] and lead to novel virus discovery such as a rhabdovirus causing haemorrhagic fever in central Africa [19]. It also provides genomic information for typing and surveillance. Real-time genomic surveillance was facilitated on-site by the portable Oxford Nanopore MinION sequencer during the 2014–16 Ebola virus (EBOV) epidemic in West Africa and the 2015-16 ZIKV outbreak in the Americas [20-23] for epidemiological and transmission chain investigations [24]. In both examples, an amplicon sequencing approach was used, but viruses and bacteria from clinical, environmental and vector samples have been sequenced using metagenomic approaches on the MinION [25-28]. Metagenomic sequencing of CHIKV was demonstrated in principle on the MinION by Greninger et al. in 2015 reporting the detection of CHIKV from a human blood sample [28]. Additionally, Illumina-based metagenomics identified CHIKV coinfections within a ZIKV sample cohort [29], with the high proportion of CHIKV reads present making it a promising target for the approach.

In this study we set out to test the feasibility of direct metagenomic sequencing of DENV and CHIKV genomes from a cohort of clinical serum and plasma samples across a representative range of viral loads. The objective was to assess the proportion of viral nucleic acid relative to patient/background present in each sample and determine the sequencing limits for whole genome retrieval using both the laboratory-based Illumina technology and the portable MinION platform.

Methods

Sample collection and nucleic acid extraction

Twenty-six routine diagnostic samples, nine plasma and 17 serum, were obtained from the Rare and Imported Pathogens Laboratory (RIPL), Public Health England (PHE), Porton Down. All had previously tested positive by real-time reverse transcription-PCR (qRT-PCR) for chikungunya or dengue virus, with a maximum cut-off value of cycle threshold (Ct) 35. These samples had been selected based on their Ct values, among a larger set of 441 samples, so as to represent a Ct clinical range. Total nucleic acid was extracted from 140 µL of each using the QIAamp viral RNA kit (Qiagen, Hilden, Germany) replacing carrier RNA with linear polyacrylamide and eluting in 60 µL elution buffer provided in the kit, followed by treatment with TURBO DNase (Thermo Fisher Scientific, Waltham, United States (US)) at 37 °C for 30 min. RNA was purified and concentrated to 8 µL using the RNA Clean and Concentrator-5 kit (Zymo Research, Irvine, US).

Molecular confirmation and quantification

Drosten et al. [30] and Edwards et al. [31] RT-PCR assays were used for confirmation of DENV and CHIKV respectively. RNA oligomers were used as standards for genome copy quantitation.

Table 1

Description of samples positive for chikungunya and dengue virus by real-time reverse transcription-PCR with corresponding Illumina mapping data, United Kingdom, 2017[a] (n = 26 samples)

Sample	Ct value	Estimated genome copy number in the sample (/mL)	Sample type	Total reads (R1 + R2)[b]	% reads mapping to reference viral genome	% 20x coverage	% 10x coverage	Reference virus[c]	Reference size (nts)
CHIKV 1	14.72	2.12E+10	Plasma	1,113,560	78.32	99.59	99.72	CHIKV	11,826
CHIKV 2	20.06	5.49E+08	Serum	1,278,624	98.48	99.14	99.47	CHIKV	11,826
CHIKV 3	21.41	2.18E+08	Plasma	1,391,258	95.23	98.86	99.37	CHIKV	11,826
CHIKV 4	22.74	8.76E+07	Plasma	888,968	19.16	97.08	97.32	CHIKV	11,826
CHIKV 5	23.77	4.33E+07	Plasma	1,357,606	97.13	99.16	99.58	CHIKV	11,826
CHIKV 6	25.4	1.42E+07	Serum	3,236,848	34.88	97.80	98.40	CHIKV	11,826
CHIKV 7	25.76	1.11E+07	Plasma	3,748,070	72.77	99.04	99.56	CHIKV	11,826
CHIKV 8	28.17	2.13E+06	Plasma	1,499,952	28.41	98.69	99.00	CHIKV	11,826
CHIKV 9	30.08	5.76E+05	Serum	1,035,026	6.66	95.98	98.22	CHIKV	11,826
CHIKV 10	30.37	4.72E+05	Serum	1,575,222	16.84	97.39	98.01	CHIKV	11,826
CHIKV 11	30.66	3.87E+05	Serum	1,143,054	13.52	95.36	96.96	CHIKV	11,826
CHIKV 12	30.95	3.17E+05	Serum	1,507,380	10.93	96.11	96.52	CHIKV	11,826
CHIKV 13	32.2	1.35E+05	Serum	1,323,920	5.03	88.47	89.38	CHIKV	11,826
CHIKV 14	32.57	1.05E+05	Serum	1,479,404	21.72	96.32	96.93	CHIKV	11,826
DENV 1	16.29	4.21E+09	Plasma	439,292	93.44	99.51	99.58	DENV 1	10,735
DENV 2	16.85	2.83E+09	Serum	513,472	92.56	99.40	99.58	DENV 1	10,735
DENV 3	17.67	1.58E+09	Plasma	738,814	92.53	99.58	99.58	DENV 2	10,723
DENV 4	18.20	1.09E+09	Serum	477,368	93.97	98.73	99.12	DENV 2	10,723
DENV 5	19.73	3.67E+08	Serum	915554	89.65	99.14	99.40	DENV 2	10,723
DENV 6	21.22	3.61E+07	Serum	3,587,926	83.87	99.68	99.69	DENV 4	10,649
DENV 7	23.76	2.11E+07	Serum	4,146,678	2.17	86.99	89.13	DENV 1	10,735
DENV 8	24.8	1.01E+07	Serum	777,264	69.23	99.56	99.58	DENV 3	10,707
DENV 9	25.28	7.17E+06	Plasma	787,728	26.97	98.77	98.81	DENV 2	10,723
DENV 10	26.98	2.15E+06	Serum	596,240	6.58	93.47	93.97	DENV 3	10,707
DENV 11	28.69	6.39E+05	Serum	1,034,698	3.73	94.44	94.70	DENV 1	10,735
DENV 12	31.29	1.01E+05	Serum	1,374,766	0.47	71.46	77.76	DENV 1	10,735

CHIKV: chikungunya virus; Ct: cycle threshold; DENV: dengue virus.

[a] The Illumina mapping data presented in the table were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

[b] 'R1 + R2' indicates paired-end sequencing.

[c] For DENV the serotype is also indicated.

Metagenomic cDNA reparation

Complementary DNA (cDNA) was prepared using a Sequence Independent Single Primer Amplification (SISPA) approach adapted from Greninger et al. [28]. Reverse transcription and second strand cDNA synthesis were as described [28]. cDNA amplification was performed using AccuTaq LA (Sigma, Poole, United Kingdom), in which 5µL of cDNA and 1 µL (100 pmol/µL) Primer B (5'-GTTTCCCACTGGAGGATA-3') were added to a 50 µL reaction, according to manufacturer's instructions. PCR conditions were 98 °C for 30s, followed by 30 cycles of 94 °C for 15 s, 50 °C for 20 s, and 68 °C for 5 min, and a final step of 68 °C for 10 min. Amplified cDNA was purified using a 1:1 ratio of AMPure XP beads (Beckman Coulter, Brea, California (CA)) and quantified using the Qubit High Sensitivity dsDNA kit (Thermo Fisher, Waltham, US).

MinION library preparation and sequencing

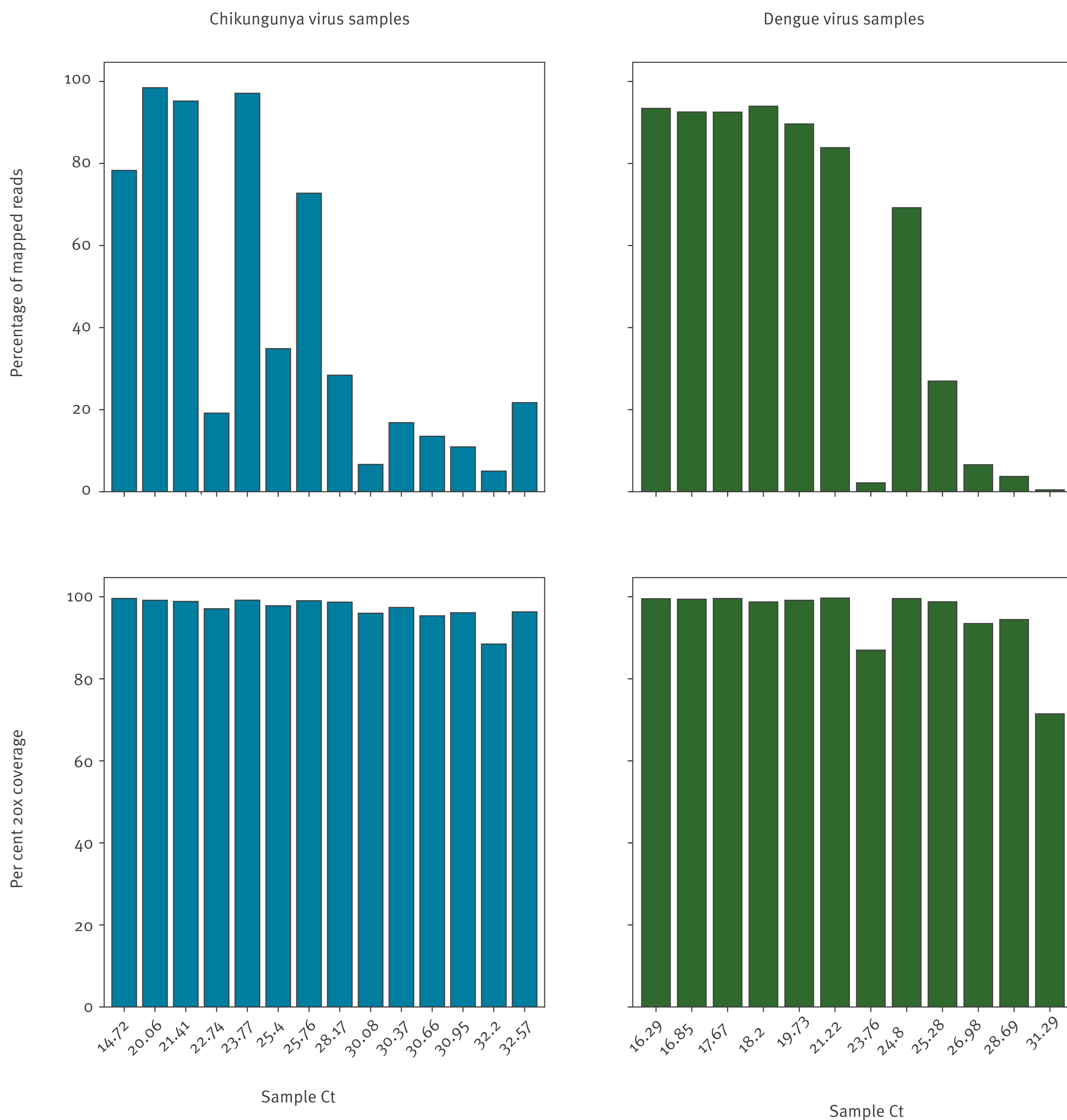
MinION sequencing libraries were prepared using total amplified cDNA of each sample to a maximum of 1 µg. Oxford Nanopore kits SQK-NSK007 or SQK-LSK208 (2D), SQK-LSK308 (1D^2) and SQK-RBK001 (Rapid) were used and each sample was run individually on the appropriate flow cell (FLO-MIN105, FLO-MIN106 or FLO-MIN107) using the 48hr run script. Base calling was performed using Metrichor (ONT) for SQK-NSK007 and SQK-LSK208 or Albacore v1.2 for SQK-LSK308 and SQK-RBK001. Poretools [32] was used to extract FASTQ files from Metrichor FAST5 files.

Illumina library preparation and sequencing

Nextera XT V2 kit (Illumina) sequencing libraries were prepared using 1.5 ng of amplified cDNA as per manufacturer's instructions. Samples were multiplexed in batches of a maximum of 16 samples per run and

FIGURE 2

Proportion of reads mapping to the appropriate viral reference sequence and proportion of reference genome sequenced at minimum 20-fold coverage in each chikungunya or dengue virus positive sample, United Kingdom, 2017[a] (n = 26 samples)



Ct: cycle threshold.

[a] The Illumina data presented in the figure were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

The percentage of total reads mapping to the appropriate reference sequence is plotted in the upper panel. Lower panels display the percentage of the reference genome sequenced to a minimum depth of 20-fold in the Illumina data.

Table 2

Description of chikungunya and dengue virus positive samples by real-time reverse transcription-PCR and corresponding Nanopore sequencing data, United Kingdom, 2017[a] (n = 8 samples)

Sample	Ct value	cDNA amount used for the library (ng)	Sequencing kit (2D kit version)	Flow cell (FLO-)	1D total bp	1D total reads	1D mean read length (nt)	1D max read length (nt)
CHIKV 1	14.7	431.5	SQK-NSK007	MIN105	1.51E+08	267,171	564	92,712
CHIKV 3	21.4	928.8	SQK-LSK208	MIN106	1.63E+09	1,891,028	862	99,031
CHIKV 4	22.7	113.4	SQK-NSK007	MIN105	1.74E+08	216,493	805	125,387
CHIKV 9	30.1	212.4	SQK-LSK208	MIN106	2.12E+09	3,481,358	608	121,711
DENV 1	16.3	1,626.0	SQK-NSK007	MIN105	2.42E+08	284,622	851	115,494
DENV 2	16.9	1,626.0	SQK-NSK007	MIN105	1.55E+08	203,700	760	52,157
DENV 6	21.2	475.0	SQK-LSK208	MIN106	1.22E+09	1,377,721	886	118,733
DENV 11	28.7	65.8	SQK-LSK208	MIN106	7.07E+08	1,111,566	636	119,438

Ct: cycle threshold.

[a] The Nanopore data presented in the table were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

sequenced on a 2x150 bp-paired end Illumina MiSeq run, by Genomics Services Development Unit, PHE.

Data handling

BWA MEM v0.7.15 [33] was used to align reads to the following references (GenBank ID): DENV Serotype 1 (NC_001477.1), DENV Serotype 2 (NC_001474.2), DENV Serotype 3 (NC_001475.2), DENV Serotype 4 (NC_002640.1) and CHIKV (NC_004162.2) using -x ont2d mode for Nanopore and MEM defaults for Illumina reads. Samtools v1.4 [34] was used to compute percentage reads mapped and coverage depth. Bedtools v2.26.0 [35] was used to calculate genome coverage at 10x and 20x. Mapping consensuses for Illumina were generated using in-house software QuasiBam [36] and for MinION using a simple pileup with bases called at a minimum depth of 20x and 70% support fraction. Nanopolish variants [24,37] was used in consensus mode to compute an error-corrected consensus sequence from the Rapid kit data. Taxonomic classification was performed using Kraken (0.10.4-beta) [38] and a locally built database populated with all RefSeq bacterial, viral, and archaeal genomes plus additional sequences [39]. De novo assemblies were generated using Spades 3.8.2 [40] in combination with SSPACE Standard v3.0 [41] for Illumina generated sequences and Canu v1.6 [41,42] for Nanopore sequences (settings: corOutCoverage = 1,000; genomeSize = 12,000; minReadLength = 300, minOverlapLength = 50).

Consensus sequences for all samples tested are available in Genbank, raw fast5 files from 1D2 and 1D data (viral reads only) are deposited in SRA (Both under BioProject PRJNA508296).

Results

Metagenomic Illumina sequencing

A total of 73 samples tested during 2016 in RIPL diagnostic laboratories, PHE Porton Down, were positive by qRT-PCR for CHIKV, and 368 were positive for DENV. Median Ct for CHIKV was 26.1, for DENV it was 26.8. For each virus, samples representing the range of viral titres seen during 2016 were selected, based on qRT-PCR Ct value (Figure 1). CHIKV samples selected (n = 14) ranged from Ct 14.72 to Ct 32.57, corresponding to 10^{10} and 10^5 genome copies per mL of plasma or serum. DENV samples selected (n = 12) ranged from Ct 16.29 to Ct 31.29, corresponding to 10^9 and 10^5 genome copies per mL (Table 1). To measure the proportion of viral nucleic acid present relative to host/background and assess genome coverage, all samples were processed as described in methods and Illumina sequenced (Table 1). The proportion of total reads mapping to the respective viral reference was high for both viruses (Figure 2). In some low Ct samples, over 90% of reads mapped to the viral reference and proportions over 50% were still observed at mid-Ct range. The lowest proportions observed were 5.03% and 0.47% for CHIKV and DENV respectively (Table 1, Figure 2). The majority of samples returned over 95% genome coverage at 20x (21/26 samples) and over 98% genome coverage at 10x (20/26 samples). Irrespective of lower mapping percentages in high Ct value samples, genome coverage of 88.5% (20x) and 89.4% (10x) for CHIKV and 75.0% (20x) and 77.8% (10x) for DENV was observed.

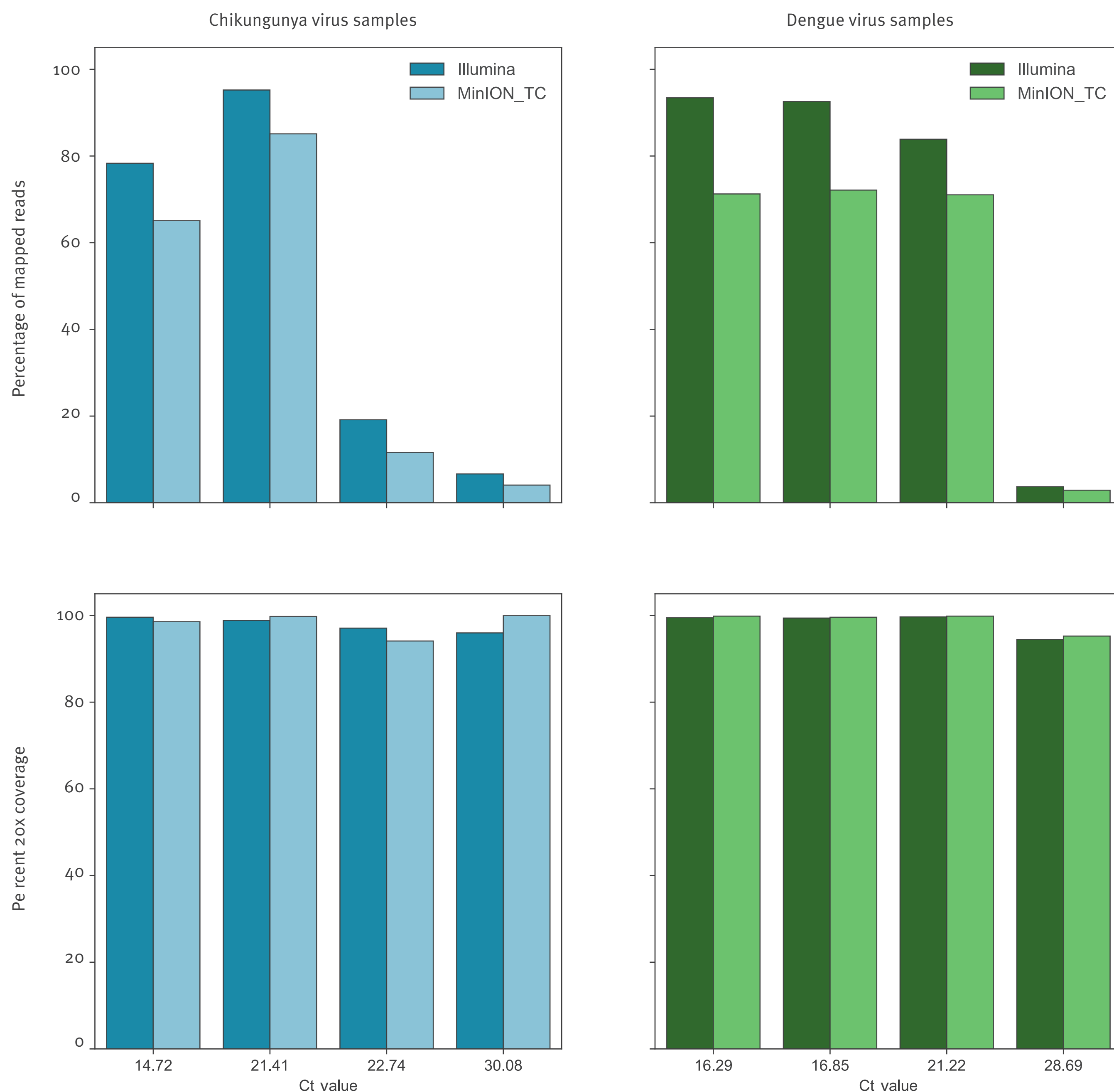
Metagenomic MinION sequencing

Four representative samples for each virus were selected for Nanopore sequencing (Table 2).

Figure 3 shows percentages of reads mapping to viral reference, which were generally concordant with the

Figure 3

Comparison of Nanopore and Illumina results, as to proportions of reads mapping to the appropriate reference viral sequence, and proportions of reference genome sequenced at minimum 20-fold coverage, United Kingdom, 2017[a] (n = 8 samples)



Ct: cycle threshold; TC: template/complement.

[a] The Nanopore and Illumina data presented in the figure were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

The percentage of total reads mapping to the appropriate reference sequence is plotted in the upper panel. Lower panels display the percentage of the reference genome sequenced to a minimum depth of 20-fold in the data generated, in dark blue or dark green for the Illumina sequence data, in light blue or light green for Nanopore data (MinION_TC).

Table 3

Summary of Nanopore mapping data on chikungunya and dengue virus positive samples by real-time reverse transcription-PCR, United Kingdom, 2017[a] (n = 8 samples)

Sample	Ct value	Total reads	% reads mapping to appropriate viral sequence	% 20x coverage	20x genome length (nt)	% 10x coverage	Reference[b]	Reference size (nt)	Max de novo contig (nt)
CHIKV 1	14.7	267,171	65.1	98.57	11,658	99.2	CHIKV	11,826	5,263
CHIKV 3	21.4	1,891,028	85.1	99.76	11,798	99.9	CHIKV	11,826	10,793
CHIKV 4	22.7	216,493	11.6	94.11	11,130	97.2	CHIKV	11,826	4,256
CHIKV 9	30.08	3,481,358	4.08	100	11,826	100	CHIKV	11,826	9,860
DENV 1	16.3	284,622	71.3	99.9	10,719	99.9	DENV 1	10,735	8,281
DENV 2	16.9	203,700	72.1	99.6	10,692	99.6	DENV 1	10,735	10,157
DENV 6	21.2	1,377,721	71.1	99.9	10,634	99.9	DENV 4	10,649	7,877
DENV 11	28.7	1,111,566	2.9	95.3	10,226	96.3	DENV 1	10,735	4,699

CHIKV: chikungunya virus; Ct: cycle threshold; DENV: dengue virus.

[a] The Nanopore data presented in the table were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

[b] For DENV the serotype is also indicated.

Illumina data, although a slight decrease is observed across the range of Ct values. In the Nanopore data, the highest mapped read percentages observed were 85.12% and 72.14% for CHIKV3 and DENV 2 respectively, compared with 95.23% and 92.56% in the Illumina data from the same samples. While in high Ct samples the viral proportion drops to 4.08% for CHIKV 9 and 2.90% for DENV 11, from 6.66% and 3.73% in the Illumina data.

Despite the decrease in proportion of mapped viral reads, comparable genome coverage is observed at both 20x and 10x (Figure 3, Table 3) and is even increased compared with Illumina data at lower viral titres, e.g. 100% at 20x for CHIKV 9 compared with 95.98% in the Illumina data and 95.25% for the high Ct DENV 11 sample, which generated 94.44% coverage from the Illumina data. Average read lengths in Nanopore data ranged from 564 to 886 bp (Table 2).

Figure 4 shows coverage depth of reads mapped across the relevant genome for each sample sequenced by both Illumina and Nanopore. Read levels are not normalised thus actual depth is a function of total reads sequenced, but the pattern of coverage seen is highly similar suggesting it is more dependent upon the SISPA methodology than sequencing library preparation. From Nanopore consensus genome sequences, between 99.93% and 100% of bases called per sample agreed with the Illumina generated sequence.

Metagenomic data analysis and coinfection identification

To test the applicability of a metagenomic analysis approach to the data, we assessed read taxonomic classification using Kraken (Figure 5). The distribution of reads classified as CHIKV, DENV, other viruses, bacteria, and archaea/eukaryota show a similar pattern for Illumina and Nanopore data. The proportion of unclassified reads for each sample increased with Ct value, as the proportion of human origin reads is higher and the human genome is not represented in our Kraken database. A decrease in the percentage of CHIKV and DENV classified reads is observed for MinION data compared with Illumina, but was sufficient to identify the correct predominant virus in all samples.

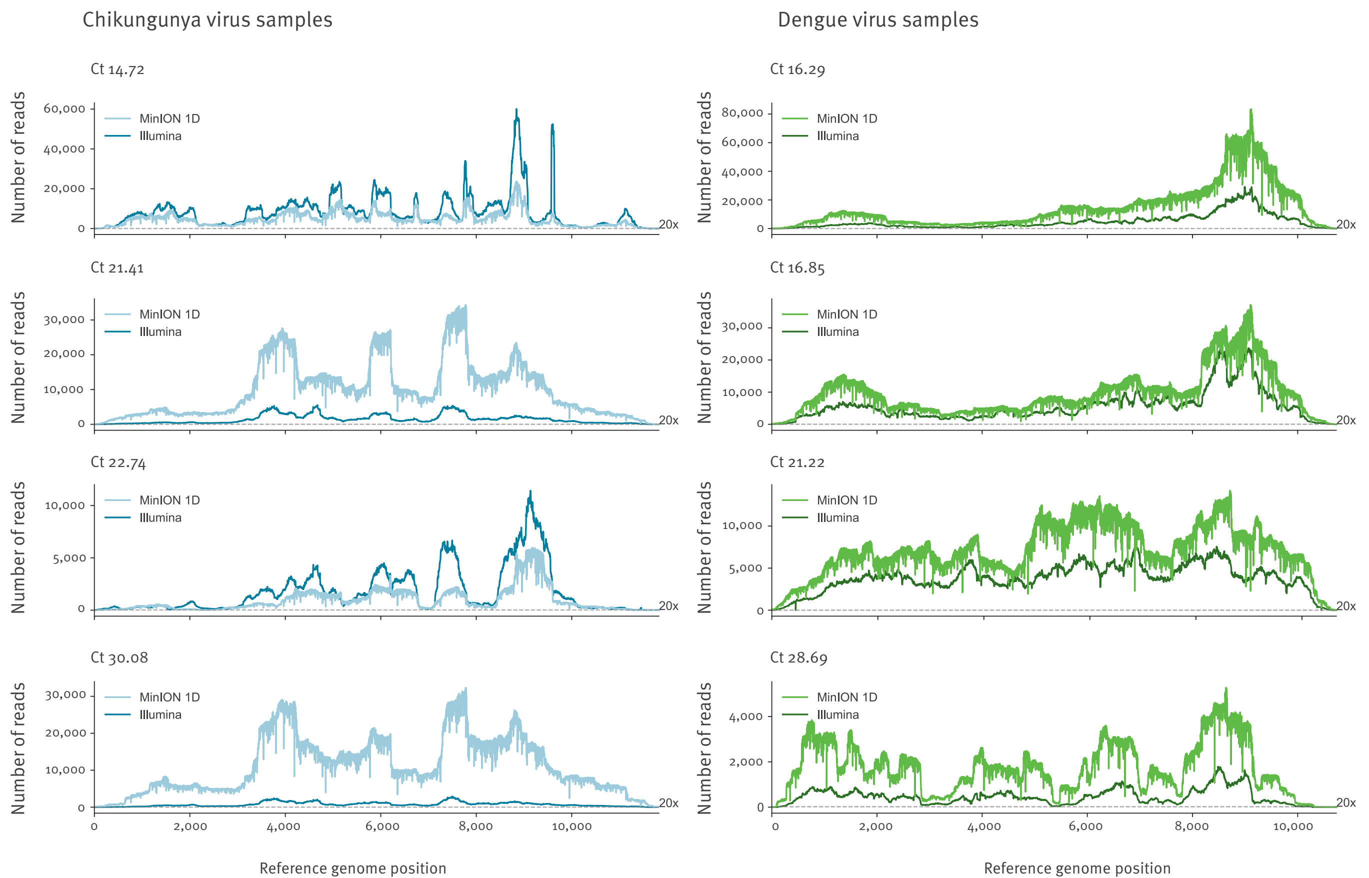
Kraken analysis also allowed for the identification of a DENV coinfection in sample CHIKV 3, the consensus sequence of which was unique in the sample set, eliminating cross-contamination from the DENV positive samples as potential source. Kraken classified 0.08% of Illumina reads and 0.15% of MinION reads as DENV. Using reference mapping to validate the finding, 0.22% of Illumina reads and 0.43% of MinION reads mapped to a DENV-1 reference genome. Genome coverage at 20x of 99.73% and 95.99% was achieved for the primary CHIKV and secondary DENV coinfection respectively, with a single MinION flow cell.

De novo assembly

De novo assembly of the data was attempted using Canu [42] and contigs identified using Basic Local Alignment Search Tool against a Nt database (BLASTn). Table 3 lists the longest viral contig length identified in each sample, ranging from 4.2 Kb (36% of reference genome size) to 10.8 Kb (91%) for CHIKV and 4.7 Kb (44%) to 10.1 Kb (95%) for DENV. Identification of the pathogen present without prior knowledge would have therefore been possible for all samples.

Figure 4

Coverage depth across the chikungunya or dengue viral genome, United Kingdom, 2017[a] (n = 8 samples)



Ct: cycle threshold.

[a] The Nanopore and Illumina data presented in the figure were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

Each graph corresponds to a given sample, defined by its Ct value. Read depth (y-axis) across the genome (x-axis) following reference alignment is shown. Illumina coverage is shown in darker blue and darker green for chikugunya and dengue virus positive samples respectively. Nanopore (MinION) coverage is indicated in lighter blue or lighter green for chikungunya and dengue virus positive samples respectively. Total depth has not been normalised; comparison is to show overall pattern of coverage is highly similar across the methods. Dotted horizontal line indicates depth of 20x coverage, used for consensus calling.

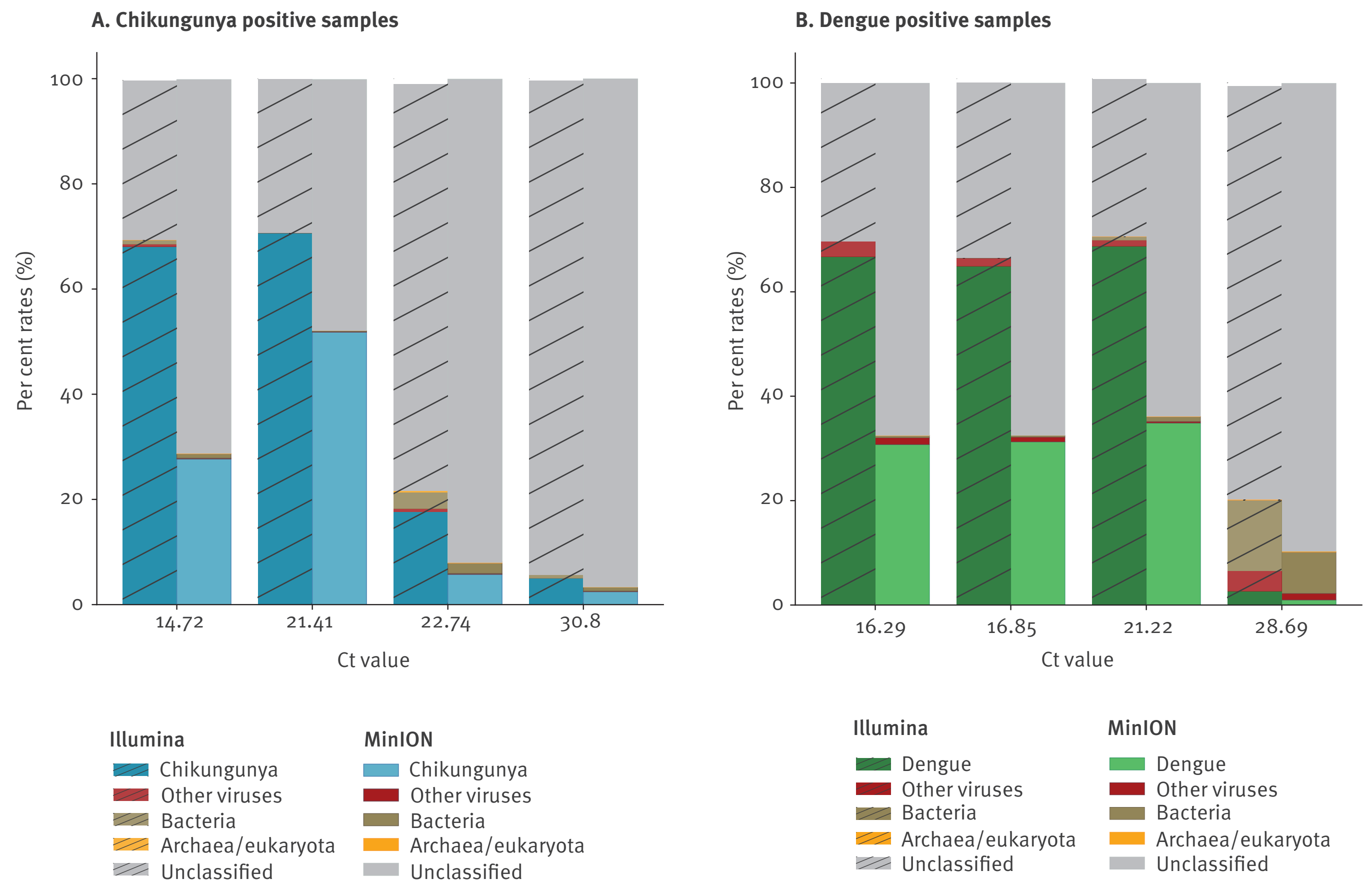
Updated MinION library kits

We repeated the sequencing of the coinfected CHIKV 3 sample using the MinION 1D² (SQK-LSK308) and Rapid (SQK-RBK001) kits, currently the most accurate and the fastest library preparation kits available, respectively. Using the 1D²kit 74.5% of reads generated mapped to CHIKV and 0.37% to DENV, while from the Rapid kit the result was 66.26% and 0.29% respectively (both lower than observed in the 2D chemistry). Coverage at 20x for CHIKV was above 99% for both kits, and for DENV was 95.04% from the 1D² and 81.09% from the Rapid kit (Table 4). Coverage depth pattern across the genome for both viruses (Figure 6) was similar for all library kits tested. Near-maximum coverage for both viruses was obtained within 30 min with the 2D kit, 8 min with the 1D² kit and 85 min with the Rapid kit (Supplementary Figure 1). De novo assembly (Table 4) produced best CHIKV contigs of 10.7, 11.3 and 11.4 Kb for the 2D, 1D² and Rapid libraries respectively and the longest contigs generated for DENV were 7.5, 2.2 and 4.2 Kb.

The 1D data from the Rapid kit was sufficient to call a consensus from 11,647/11,826 bases of the CHIKV reference with 179/11,826 bases called as ambiguous or too low coverage. All bases called were concordant with the Illumina consensus. A polishing step using Nanopolish [37] with a subset of the mapped reads (ca 100x coverage depth) significantly reduced ambiguous calls to 90/11,826, introducing a single disagreement with the Illumina consensus (99.99% concordance). Despite considerably greater read depth, the 1D² kit called only

Figure 5

Kraken classification of reads from metagenomic sequencing in (A) chikungunya and (B) dengue real-time reverse transcription-PCR positive samples, United Kingdom, 2017[a] (n = 8 samples)



Ct: cycle threshold.

[a] Read taxonomic classification using Kraken was conducted in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

Kraken classification distribution comparison for Illumina (cross-hatched) and Nanopore data. Reads grouped as either chikungunya virus (blue in panel A), dengue virus (green in panel B), other viruses (brown), archea/eukaryota (orange), bacteria (brown) or unclassified (grey).

11,082/11,826 due to a higher proportion, 744/11826, of ambiguous base calls, suggesting 1D reads are most suitable for this approach.

Discussion

These results clearly show that there are considerable levels of viral nucleic acid present in a large proportion of CHIKV and DENV qRT-PCR positive clinical samples, and demonstrate that relatively modest metagenomic sequencing is capable of elucidating significant portions of viral genome even for samples with Ct values at the higher end of clinical range. A decreased Ct value coincided with an increased proportion of viral reads, with a considerable level of variation between samples, likely because of the total level of non-viral host/background nucleic acid present due to variability between patients or in sample handling during collection, storage and testing. For example, the two lowest viral titre CHIKV samples (13 and 14) have similar Ct values (32.2 and 32.57) but varied significantly in the proportion of viral reads (5.03% and 21.72%). The 5.03% viral reads in CHIKV13 is the lowest for CHIKV, yet still sufficient to generate 88.5% of the CHIKV genome at 20x depth from just ca 662,000 paired-end Illumina reads. This amount of genomic information is highly informative and further sequencing would likely increase coverage. Only seven of the 73 total CHIKV diagnostic samples tested in 2016 had a Ct greater than 32.2 (including sample CHIKV14) (Table 1), which suggests that for the majority (>90%) of CHIKV PCR positive samples, viral load is sufficient for genome sequencing directly from patient samples without further viral enrichment beyond a simple DNAse digestion (Figure 1). The lowest viral read proportion observed in the DENV samples was 0.47% in DENV12, Ct 31.29, which generated 71.5% coverage at 20x depth (increased to 77.8 at 10x

Table 4

Comparison of Nanopore mapping data across library kits, United Kingdom, 2017[a] (n = 8 samples)

Platform	Kit information	Flow cell (FLO-)	Virus identified	Total reads (nt)	% reads mapping	% 20x coverage	% 10x coverage	Reference[b]	Reference size (nt)	Max de novo contig (nt)
Illumina	Nextera XT	NA	CHIKV	1,391,258	95.23	98.86	99.37	CHIKV	11,826	7,321
Illumina	Nextera XT	NA	DENV	1,391,258	0.22	63.66	77.82	DENV1	10,735	6,613
MinION 2D	SQK-LSK208	MIN106	CHIKV	1,891,028	85.12	99.73	99.91	CHIKV	11,826	10,793
MinION 2D	SQK-LSK208	MIN106	DENV	1,891,028	0.43	95.99	96.09	DENV1	10,735	7,549
MinION 1D^2	SQK-LSK308	MIN107	CHIKV	5,080,906	74.50	99.94	100	CHIKV	11,826	11,369
MinION 1D^2	SQK-LSK308	MIN107	DENV	5,080,906	0.37	95.04	96.42	DENV1	10,735	2,199
MinION Rapid	SQK-RBK001	MIN106	CHIKV	611,110	66.26	99.66	99.68	CHIKV	11,826	11,473
MinION Rapid	SQK-RBK001	MIN106	DENV	611,110	0.29	81.09	90.83	DENV1	10,735	4,227

CHIKV: chikungunya virus; Ct: cycle threshold; DENV: dengue virus. NA: not applicable.

[a] Results presented in the table were obtained in 2017 on samples that had been collected and found positive for chikungunya or dengue virus by real-time reverse transcription-PCR in 2016.

[b] For DENV the serotype is also indicated.

depth) from just 687,000 paired end Illumina reads and allowed for DENV serotype identification. Only 62 of 368 DENV cases in 2016 had a higher Ct, predicting that > 80% of PCR positive DENV samples have a viral load sufficient for genome sequencing (Figure 1). These estimates are based on Ct range distribution from a single year, results may vary from year to year.

The high yield of viral sequences from clinical CHIKV and DENV samples make the exciting prospect of metagenomic MinION viral whole-genome-sequencing feasible, even for lower viral titre samples. Evaluating this on a representative subset of our samples demonstrates that viral read proportions are in general agreement with that seen for Illumina sequencing, predicting a similar proportion of qRT-PCR positive patient samples would be suitable for direct metagenomic sequencing on the MinION. Differences in precise proportions of viral reads seen between Illumina and MinION are likely due to inter-library variation. Differences in genome coverage achieved are due to both differences in total reads generated per sample (not normalised between platforms) as well as differences in average read length. Of the samples tested on the MinION, the lowest titre samples CHIKV 9 and DENV 11 both generated near complete genome coverage.

We repeated the sequencing of the coinfected CHIKV 3 sample using the MinION 1D^2 (SQK-LSK308) and Rapid (SQK-RBK001) kits. A reduction in viral proportion of total reads was observed compared with the 2D kit, which may be due partly to the extended storage time of the original samples before retesting. In the case of the 1D^2 kit, the lower proportion was outweighed by a substantial increase in total data generated per flow cell (5 M vs 1.8 M reads). For the Rapid kit, the total data produced should be considered in the light of the greatly simplified sample workflow and turnaround-time.

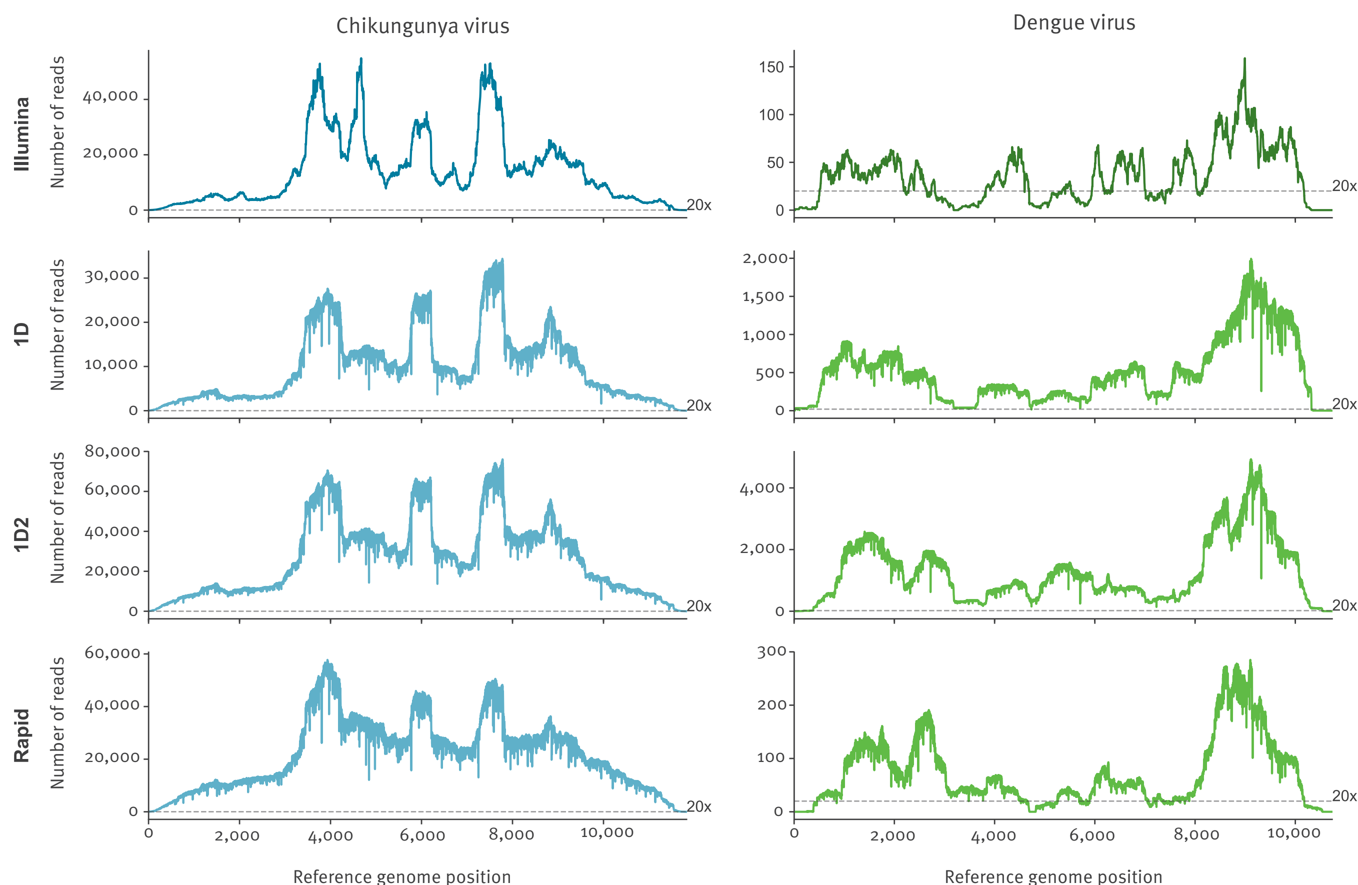
The use of metagenomics to elucidate genomic sequences of RNA viruses directly from clinical samples has several obvious benefits in public health applications. The primary benefit over targeted methods is the hypothesis-free nature of the assay, which allows identification and genomic characterisation of novel or unexpected RNA viral agents, either as primary or coinfectants (demonstrated here in the CHIKV/DENV coinfection sample), without any prior clinical knowledge. It also removes the need for laboratory optimisation of targeted methods, such as primer or bait-probe design and testing, and is not subject to escape mutations in target sites that afflict targeted sequencing and diagnostic methods. This issue particularly relevant for highly diverse RNA viruses, such as Lassa virus, which are difficult to assess using targeted methods, without regular reappraisal [43].

The principal limitation of the metagenomic approach is the limit of detection. The data generated here show that viral titres as low as 10^5 are sufficient for significant genome recovery by this method, but ZIKV is a recent example of a pathogen typically present at lower clinical titres, for which targeted methods are an absolute requirement [22,23]. For diagnostic purposes qRT-PCR has a lower limit of detection, provided the target site is conserved in the pathogen isolate tested. Clearly no single method is most suitable for both detection and genotyping of all pathogens and each has a role to play in differing circumstances.

The ability to generate genomic data directly from patient samples is clearly of great benefit to public

Figure 6

Comparison of genome coverage depth across the chikungunya virus or dengue virus genome for different sequencing library preparation methods in a sample coinfected with dengue and chikungunya viruses, United Kingdom, 2017[a] (n = 1 sample)



[a] Results presented in the Figure were obtained in 2017 on a sample that had been collected and found positive for chikungunya virus by real-time reverse transcription-PCR in 2016. In 2017, the sample was further found to be coinfected with CHIKV and DENV by metagenomic sequencing.

Read depth across both CHIKV and DENV genomes following reference alignment is shown for coinfection sample CHIKV 3, sequenced using four different sequencing library preparation/sequencing methods. Total coverage depth has not been normalised; comparison is to show overall pattern of coverage is highly similar across the methods. Dotted horizontal line indicates depth of 20x coverage, used for consensus calling.

health (reviewed in detail [44]). It can be used in a routine surveillance capacity or early during suspected outbreaks to link related cases who may be missed by traditional epidemiology [45] and identify outbreak cases distinct from typically circulating seasonal strains, which is key in regions endemic for the pathogen in question. The use of whole genome sequences offers the greatest precision for these applications, compared with typing methods based on specific genomic regions [44]. Whole genome sequencing on a portable device allows this information to be generated rapidly and within the affected region [24], enabling timely identification of an outbreak, or allaying fears of a potential one if cases are not linked. Furthermore mutations relating to viral drug resistance or pathogenicity can be monitored [44]. Therefore the ability to generate near-complete viral genome sequences directly from clinical samples on a portable sequencing device has many potential applications.

Conclusions

We demonstrate that across the clinically relevant range of viral loads an unexpectedly high proportion of reads generated metagenomically from CHIKV and DENV clinical samples are viral in origin. Therefore metagenomic sequencing provides an effective approach for the analysis of CHIKV and DENV genomes directly from the majority of qRT-PCR positive serum and plasma samples, without the need for culture or viral nucleic acid enrichment beyond a simple DNA digestion. We demonstrate this is equally possible on the Oxford Nanopore MinION, making metagenomic

whole genome sequencing potentially feasible in the field.

Acknowledgements

This work was funded via an NIHR HPRU in Emerging and Zoonotic Infections PhD studentship awarded to L. Kafetzopoulou. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health. Oxford Nanopore Technologies provided some reagents free of charge and funded author conference attendance.

Conflict of interest

Oxford Nanopore Technologies provided some reagents free of charge and funded author conference attendance.

Authors' contributions

Performed experiments: LEK, KL, AC, DC

Performed Data analysis: LEK, KE, STP

Design of study: LEK, EA, JO, RH, JAH, MWC, RV, STP

Wrote the manuscript: LEK, STP

All authors reviewed the manuscript.

References

1. Papa A. Emerging arboviral human diseases in Southern Europe. J Med Virol. 2017;89(8):1315-22. https://doi.org/10.1002/jmv.24803 PMID: 28252204
2. Gould E, Pettersson J, Higgs S, Charrel R, de Lamballerie X. Emerging arboviruses: Why today? One Health. 2017;4:1-13. https://doi.org/10.1016/j.onehlt.2017.06.001 PMID: 28785601
3. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85(2):328-45. https://doi.org/10.1016/j.antiviral.2009.10.008 PMID: 19857523
4. Thiberville S-D, Moyen N, Dupuis-Maguiraga L, Nougairede A, Gould EA, Roques P, et al. Chikungunya fever: epidemiology, clinical syndrome, pathogenesis and therapy. Antiviral Res. 2013;99(3):345-70. https://doi.org/10.1016/j.antiviral.2013.06.009 PMID: 23811281
5. World Health Organization (WHO). WHO publishes list of top emerging diseases likely to cause major epidemics. Geneva:WHO; 2017 Nov 10. [Accessed 26 Feb 2018]; Available from: http://www.who.int/medicines/ebola-treatment/WHO-list-of-top-emerging-diseases/en/
6. Burt FJ, Chen W, Miner JJ, Lenschow DJ, Merits A, Schnettler E, et al. Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. Lancet Infect Dis. 2017;17(4):e107-17. https://doi.org/10.1016/S1473-3099(16)30385-1 PMID: 28159534
7. Mavalankar D, Shastri P, Bandyopadhyay T, Parmar J, Ramani KV. Increased mortality rate associated with chikungunya epidemic, Ahmedabad, India. Emerg Infect Dis. 2008;14(3):412-5. https://doi.org/10.3201/eid1403.070720 PMID: 18325255
8. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1545-602. https://doi.org/10.1016/S0140-6736(16)31678-6 PMID: 27733282
9. Patterson J, Sammon M, Garg M. Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World. West J Emerg Med. 2016;17(6):671-9. https://doi.org/10.5811/westjem.2016.9.30904 PMID: 27833670
10. Burt FJ, Rolph MS, Rulli NE, Mahalingam S, Heise MT. Chikungunya: a re-emerging virus. Lancet. 2012;379(9816):662-71. https://doi.org/10.1016/S0140-6736(11)60281-X PMID: 22100854
11. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue. N Engl J Med. 2012;366(15):1423-32. https://doi.org/10.1056/NEJMra1110265 PMID: 22494122
12. Furuya-Kanamori L, Liang S, Milinovich G, Soares Magalhaes RJ, Clements ACA, Hu W, et al. Co-distribution and co-infection of chikungunya and dengue viruses. BMC Infect Dis. 2016;16(1):84. https://doi.org/10.1186/s12879-016-1417-2 PMID: 26936191
13. Perera-Lecoin M, Luplertlop N, Surasombatpattana P, Liégeois F, Hamel R, Thongrungkiat S, et al. Dengue and Chikungunya Coinfection – The Emergence of an Underestimated Threat. In: Rodriguez-Morales AJ, editor. Current Topics in Chikungunya. InTech; 2016.
14. Omarjee R, Prat C, Flusin O, Boucau S, Tenebray B, Merle O, et al. Importance of case definition to monitor ongoing outbreak of chikungunya virus on a background of actively circulating dengue virus, St Martin, December 2013 to January 2014. Euro Surveill. 2014;19(13):20753. https://doi.org/10.2807/1560-7917.ES2014.19.13.20753 PMID: 24721537
15. Brito CAA, Azevedo F, Cordeiro MT, Marques ETA Jr, Franca RFO. Central and peripheral nervous system involvement caused by Zika and chikungunya coinfection. PLoS Negl Trop Dis. 2017;11(7):e0005583. https://doi.org/10.1371/journal.pntd.0005583 PMID: 28704365
16. Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. Lancet Infect Dis. 2017;17(3):e101-6. https://doi.org/10.1016/S1473-3099(16)30518-7 PMID: 28011234
17. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS Pathog. 2009;5(5):e1000455. https://doi.org/10.1371/journal.ppat.1000455 PMID: 19478873
18. Towner JS, Sealy TK, Khristova ML, Albariño CG, Conlan S, Reeder SA, et al. Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Pathog. 2008;4(11):e1000212. https://doi.org/10.1371/journal.ppat.1000212 PMID: 19023410
19. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS Pathog. 2012;8(9):e1002924. https://doi.org/10.1371/journal.ppat.1002924 PMID: 23028323
20. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. Nature. 2017;546(7658):411-5. https://doi.org/10.1038/nature22402 PMID: 28538734
21. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546(7658):401-5. https://doi.org/10.1038/nature22400 PMID: 28538723
22. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546(7658):406-10. https://doi.org/10.1038/nature22401 PMID: 28538727
23. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261-76. https://doi.org/10.1038/nprot.2017.066 PMID: 28538739
24. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32. https://doi.org/10.1038/nature16996 PMID: 26840485
25. Walter MC, Zwirglmaier K, Vette P, Holowachuk SA, Stoecker K, Genzel GH, et al. MinION as part of a biomedical rapidly deployable laboratory. J Biotechnol. 2017;250:16-22. https://doi.org/10.1016/j.jbiotec.2016.12.006 PMID: 27939320
26. Brown BL, Watson M, Minot SS, Rivera MC, Franklin RB. MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach. Gigascience. 2017;6(3):1-10. https://doi.org/10.1093/gigascience/gix007 PMID: 28327976
27. Batovska J, Lynch SE, Rodoni BC, Sawbridge TI, Cogan NO. Metagenomic arbovirus detection using MinION nanopore sequencing. J Virol Methods. 2017;249:79-84. https://doi.org/10.1016/j.jviromet.2017.08.019 PMID: 28855093
28. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015;7(1):99. https://doi.org/10.1186/s13073-015-0220-9 PMID: 26416663

29. Sardi SI, Somasekar S, Naccache SN, Bandeira AC, Tauro LB, Campos GS, et al. Coinfections of Zika and Chikungunya Viruses in Bahia, Brazil, Identified by Metagenomic Next-Generation Sequencing. J Clin Microbiol. 2016;54(9):2348-53. https://doi.org/10.1128/JCM.00877-16 PMID: 27413190
30. Drosten C, Göttig S, Schilling S, Asper M, Panning M, Schmitz H, et al. Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. J Clin Microbiol. 2002;40(7):2323-30. https://doi.org/10.1128/JCM.40.7.2323-2330.2002 PMID: 12089242
31. Edwards CJ, Welch SR, Chamberlain J, Hewson R, Tolley H, Cane PA, et al. Molecular diagnosis and analysis of Chikungunya virus. J Clin Virol. 2007;39(4):271-5. https://doi.org/10.1016/j.jcv.2007.05.008 PMID: 17627877
32. Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics. 2014;30(23):3399-401. https://doi.org/10.1093/bioinformatics/btu555 PMID: 25143291
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943
35. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278
36. Penedos AR, Myers R, Hadef B, Aladin F, Brown KE. Assessment of the Utility of Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks. PLoS One. 2015;10(11):e0143081. https://doi.org/10.1371/journal.pone.0143081 PMID: 26569100
37. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12(8):733-5. https://doi.org/10.1038/nmeth.3444 PMID: 26076426
38. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46. https://doi.org/10.1186/gb-2014-15-3-r46 PMID: 24580807
39. Lewandowski K, Bell A, Miles R, Carne S, Wooldridge D, Manso C, et al. The Effect of Nucleic Acid Extraction Platforms and Sample Storage on the Integrity of Viral RNA for Use in Whole Genome Sequencing. J Mol Diagn. 2017;19(2):303-12. https://doi.org/10.1016/j.jmoldx.2016.10.005 PMID: 28041870
40. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599
41. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578-9. https://doi.org/10.1093/bioinformatics/btq683 PMID: 21149342
42. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722-36. https://doi.org/10.1101/gr.215087.116 PMID: 28298431
43. Andersen KG, Shapiro BJ, Matranga CB, Sealfon R, Lin AE, Moses LM, et al. Viral Hemorrhagic Fever Consortium. Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. Cell. 2015;162(4):738-50. https://doi.org/10.1016/j.cell.2015.07.020 PMID: 26276630
44. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol. 2017;15(3):183-92. https://doi.org/10.1038/nrmicro.2016.182 PMID: 28090077
45. Keita M, Duraffour S, Loman NJ, Rambaut A, Diallo B, Magassouba N, et al. Unusual Ebola Virus Chain of Transmission, Conakry, Guinea, 2014-2015. Emerg Infect Dis. 2016;22(12):2149-52. https://doi.org/10.3201/eid2212.160847 PMID: 27869596

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

RESEARCH

Prospective genomic surveillance of meticillin-resistant *Staphylococcus aureus* (MRSA) associated with bloodstream infection, England, 1 October 2012 to 30 September 2013

Michelle S Toleman[1,2,3], Sandra Reuter[4], Dorota Jamrozy[2], Hayley J Wilson[1], Beth Blane[1], Ewan M Harrison[1,2], Francesc Coll[5], Russell J Hope[6], Angela Kearns[6], Julian Parkhill[2], Sharon J Peacock[1,2,3,5], M Estée Török[1,3,7]

1. University of Cambridge, Department of Medicine, Cambridge, United Kingdom
2. Wellcome Sanger Institute, Hinxton, United Kingdom
3. Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
4. University of Freiburg, Institute for Infection Prevention and Hospital Epidemiology, Freiburg, Germany
5. London School of Hygiene and Tropical Medicine, London, United Kingdom
6. Public Health England, National Infection Service, Colindale, London, United Kingdom
7. Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, United Kingdom

Correspondence: Michelle Toleman (mst39@cam.ac.uk)

Citation style for this article:
Toleman Michelle S, Reuter Sandra, Jamrozy Dorota, Wilson Hayley J, Blane Beth, Harrison Ewan M, Coll Francesc, Hope Russell J, Kearns Angela, Parkhill Julian, Peacock Sharon J, Török M Estée. Prospective genomic surveillance of methicillin-resistant Staphylococcus aureus (MRSA) associated with bloodstream infection, England, 1 October 2012 to 30 September 2013. Euro Surveill. 2019;24(4):pii=1800215. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800215

Article submitted on 25 Apr 2018 / accepted on 01 Oct 2018 / published on 24 Jan 2019

Background: Mandatory reporting of meticillin-resistant *Staphylococcus aureus* (MRSA) bloodstream infections (BSI) has occurred in England for over 15 years. Epidemiological information is recorded, but routine collection of isolates for characterisation has not been routinely undertaken. Ongoing developments in whole-genome sequencing (WGS) have demonstrated its value in outbreak investigations and for determining the spread of antimicrobial resistance and bacterial population structure. Benefits of adding genomics to routine epidemiological MRSA surveillance are unknown. **Aim:** To determine feasibility and potential utility of adding genomics to epidemiological surveillance of MRSA. **Methods:** We conducted an epidemiological and genomic survey of MRSA BSI in England over a 1-year period (1 October 2012–30 September 2013). **Results:** During the study period, 903 cases of MRSA BSI were reported; 425 isolates were available for sequencing of which, 276 (65%) were clonal complex (CC) 22. Addition of 64 MRSA genomes from published outbreak investigations showed that the study genomes could provide context for outbreak isolates and supported cluster identification. Comparison to other MRSA genome collections demonstrated variation in clonal diversity achieved through different sampling strategies and identified potentially high-risk clones e.g. USA300 and local expansion of CC5 MRSA in South West England. **Conclusions:** We demonstrate the potential utility of combined epidemiological and genomic MRSA BSI surveillance to determine the national population structure of MRSA, contextualise previous MRSA outbreaks, and detect potentially high-risk lineages. These findings support the integration of epidemiological and genomic surveillance for MRSA BSI as a step towards a comprehensive surveillance programme in England.

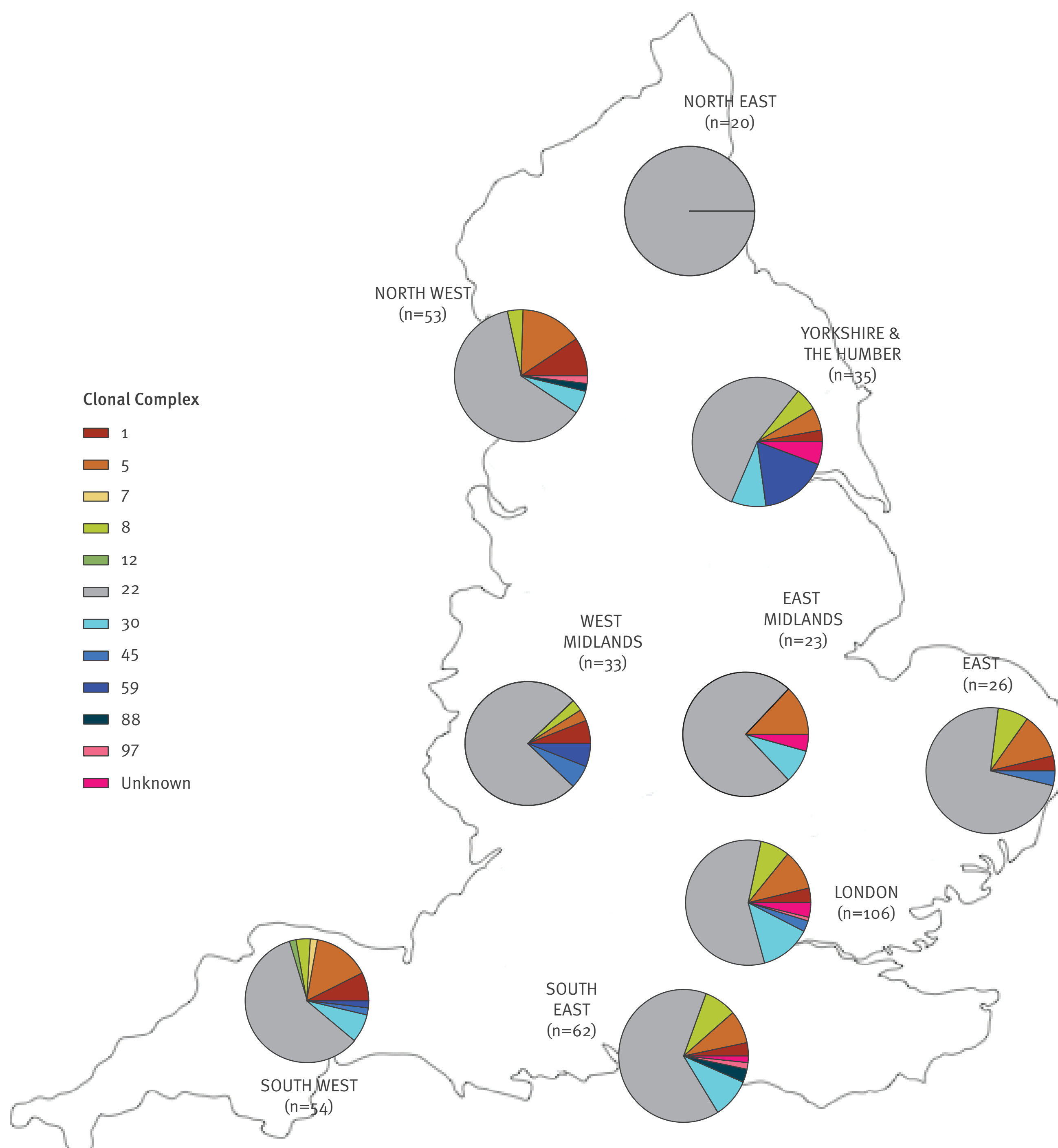
Introduction

In 2001, faced with increasingly high rates of meticillin-resistant *Staphylococcus aureus* (MRSA) bloodstream infections (BSI) at the turn of the century, the United Kingdom (UK) Department of Health mandated surveillance of MRSA BSI in England. This was followed in 2005 by enhanced surveillance to collect clinical and epidemiological information [1,2]. A number of infection prevention and control (IPC) measures were also introduced, such as strengthened antimicrobial stewardship, MRSA screening with decolonisation of all emergency hospital admissions [3] and use of care bundles for patients with intravascular catheters and indwelling urinary catheters [4]. Over the past decade, there has been a remarkable decline in the incidence of MRSA BSI in England [1,2]. Surveillance and IPC interventions are likely to have played a major role in this success, although it is unclear whether potential changes in the epidemiology of MRSA may also have contributed [5,6].

The national MRSA BSI surveillance programme conducted by Public Health England (PHE) did not include routine submission of isolates for characterisation. Isolates submitted to the PHE Staphylococcal

Figure 1

Map with breakdown of the proportions of each CC within the sequenced PHE bloodstream infection isolate collection from submitting regions, England, 1 October 2012–30 September 2013 (n = 425)



CC: clonal complex; PHE: Public Health England.

The number of isolates received by Public Health England from each of the labelled regions is indicated in brackets. Thirteen isolates did not have a region assigned.

Reference Service were highly selected and were submitted in order to type isolates for the investigation of suspected nosocomial and community outbreaks, for selected sentinel surveillance programmes and/or to detect specific genes in isolates from patients with suspected toxin-mediated disease. Strain characterisation was undertaken using staphylococcal protein A (*spa*-) typing, multilocus sequence typing (MLST), SCC*mec*-subtyping and toxin gene profiling. It is possible that a large amount of information regarding the population of disease-causing MRSA in England may have been missed as a result of this *ad hoc* approach.

Microbial WGS provides increased discriminatory power to resolve outbreaks and identify emerging MRSA lineages compared with conventional typing methods [7]. WGS has been used to investigate MRSA outbreaks in the UK [8-12] and to examine the population genetic structure of MRSA in the UK and globally [13,14]. These studies have largely been conducted through grant-funded academic research, rather than by public health programmes, using isolates from surveillance programmes such as the British Society of Antimicrobial Chemotherapy (BSAC) Antibiotic Resistance Surveillance Programme [13] or the European Antimicrobial Resistance Surveillance Network, EARS-Net [15]. Both programmes systematically collect a subset of bloodstream isolates from sentinel laboratories and routinely undertake phenotypic typing methods. Both programmes were established to monitor antimicrobial resistance, before the widespread use of WGS.

Combined with comprehensive, systematic sampling regimens WGS technologies now provide the opportunity to study the natural history of successful MRSA clones at great resolution and to identify clonal expansions to monitor in case of widespread dissemination [16]. National BSI surveillance was originally introduced in England to compare MRSA rates between hospitals and later enhanced to aid direction of clinical interventions [2]. We conducted a proof-of-principle study to determine the feasibility and potential benefits of combining prospective epidemiological and genomic surveillance of MRSA BSI on a national scale within a public health organisation. We aimed to determine what information could be gathered by combining epidemiological surveillance and routine whole-genome sequencing of isolates and to identify the potential obstacles to implementation of this strategy.

Methods

Study design, setting and participants

We conducted a prospective, observational cohort study of all cases of MRSA BSI in England from 1 October 2012 to 30 September 2013. Cases were defined as those patients reported to PHE as having a blood culture positive for MRSA by the submitting laboratory. At the time of the study the population of England, served by PHE, was approximately 53.4 million.

Data sources

In accordance with national policy, epidemiological and microbiological data on MRSA BSI cases is submitted electronically to the mandatory enhanced surveillance scheme (MESS) by infection control teams in acute National Health Service (NHS) Trusts. Mandatory data variables included patient demographics, details of hospital admission, date of BSI and location of acquisition (community or hospital). Epidemiological and microbiological data of cases with BSI during the study period was extracted from this database for use in this study. PHE reference laboratory test results were initially linked with demographic, clinical and geographic information from the MESS and then anonymised by PHE staff. Anonymised data were subsequently linked to DNA sequence data by University of Cambridge staff.

Isolate collection and laboratory testing

During the study period, all NHS diagnostic microbiology laboratories in England were invited to submit MRSA bloodstream isolates to the Staphylococcal Reference Laboratory, PHE Colindale, for characterisation. Isolates were cultured on nutrient agar and underwent *spa*-typing [17] and multiplex PCR to confirm species identification and determination of the *mecA* and *luk-PV* status [18]. Isolates were stored at -80 °C using Microbank cryovials (Pro-Laboratory Diagnostics, Cheshire, UK) pending further analyses.

DNA extraction and whole genome sequencing

Isolates were retrieved from storage, sub-cultured onto nutrient agar slopes, and transferred to the Department of Medicine at the University of Cambridge. Each sample was cultured onto Columbia Blood Agar (Oxoid, Basingstoke, UK) and identified using a commercial latex agglutination kit (Pastorex Staph Plus, Bio Rad Laboratories, Hemel Hempstead, UK). Antimicrobial susceptibility testing was performed using the Vitek-2 system (bioMérieux, Marcy l'Etoile, France). DNA was extracted, libraries prepared, and 150-bp paired-end sequences determined on an Illumina HiSeq2000 as previously described [19]. Phylogenetic trees were visualised using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and iTOL (http://itol.embl.de/).

Ethical statement

Written informed consent was not required for this study as data and isolates were collected as part of national surveillance programme for MRSA bloodstream infections, which is exempt from this requirement.

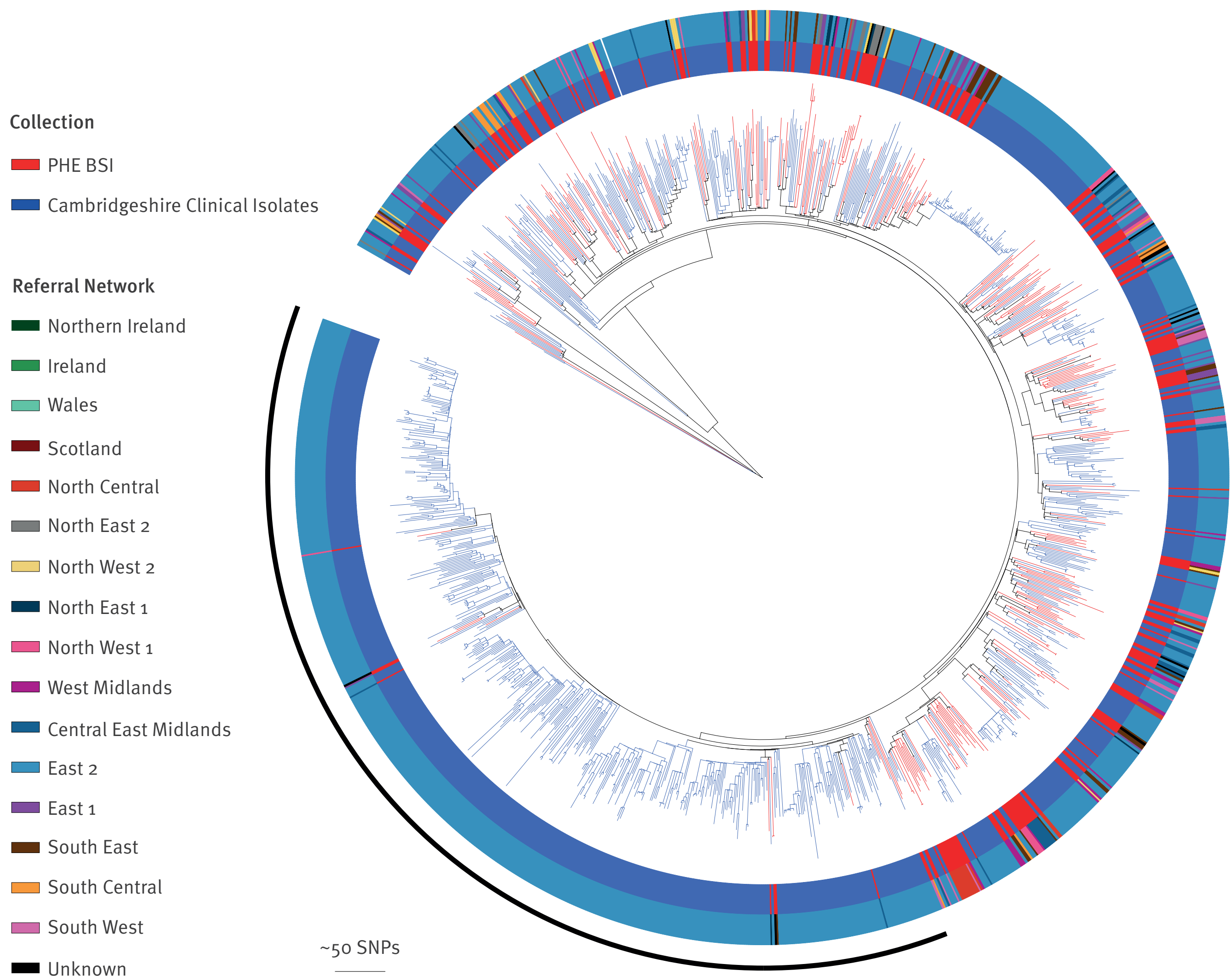
The study protocol was approved by the National Research Ethics Service (ref: 11/EE/0499), and by the Cambridge University Hospitals NHS Foundation Trust R&D Department (ref: A092428).

Genomic analysis

Genomes were assembled using an assembly and improvement pipeline [20]. MLST sequence types (STs) were assigned from the sequence data [21] (https://github.com/sanger-pathogens/mlst_check) and STs were assigned to clonal complexes (CC). Sequence data were mapped using SMALT (http://www.sanger.ac.uk/science/tools/smalt-0) to the reference genome for particular CCs (CC5, N315, GenBank *accession* number BA000018; CC8, FPR3757, GenBank *accession* number CP000255; CC22, EMRSA15, GenBank *accession* number HE681097). The core genome alignment excluding mobile genetic elements, indels and repetitive regions was generated for each CC and was used in phylogenetic estimates using RAxML with 100 bootstraps [22].

Figure 2

Phylogenetic tree comparing CC 22 PHE BSI isolates, England, October 2012-September 2013 (n = 276) to a single isolate per patient from the previously published universal sample collection from Cambridgeshire, England, April 2012–April 2013 (n = 1035)



BSI: Blood stream infections; CC: clonal complex; PHE: Public Health England; SNPs: single nucleotide polymorphisms.

Public Health England bloodstream infections CC22 isolates (n=276) are indicated by red branches. The first isolate from each patient from the previously published universal sample collection from Cambridgeshire (n=1,035) are indicated by the blue branches [24].

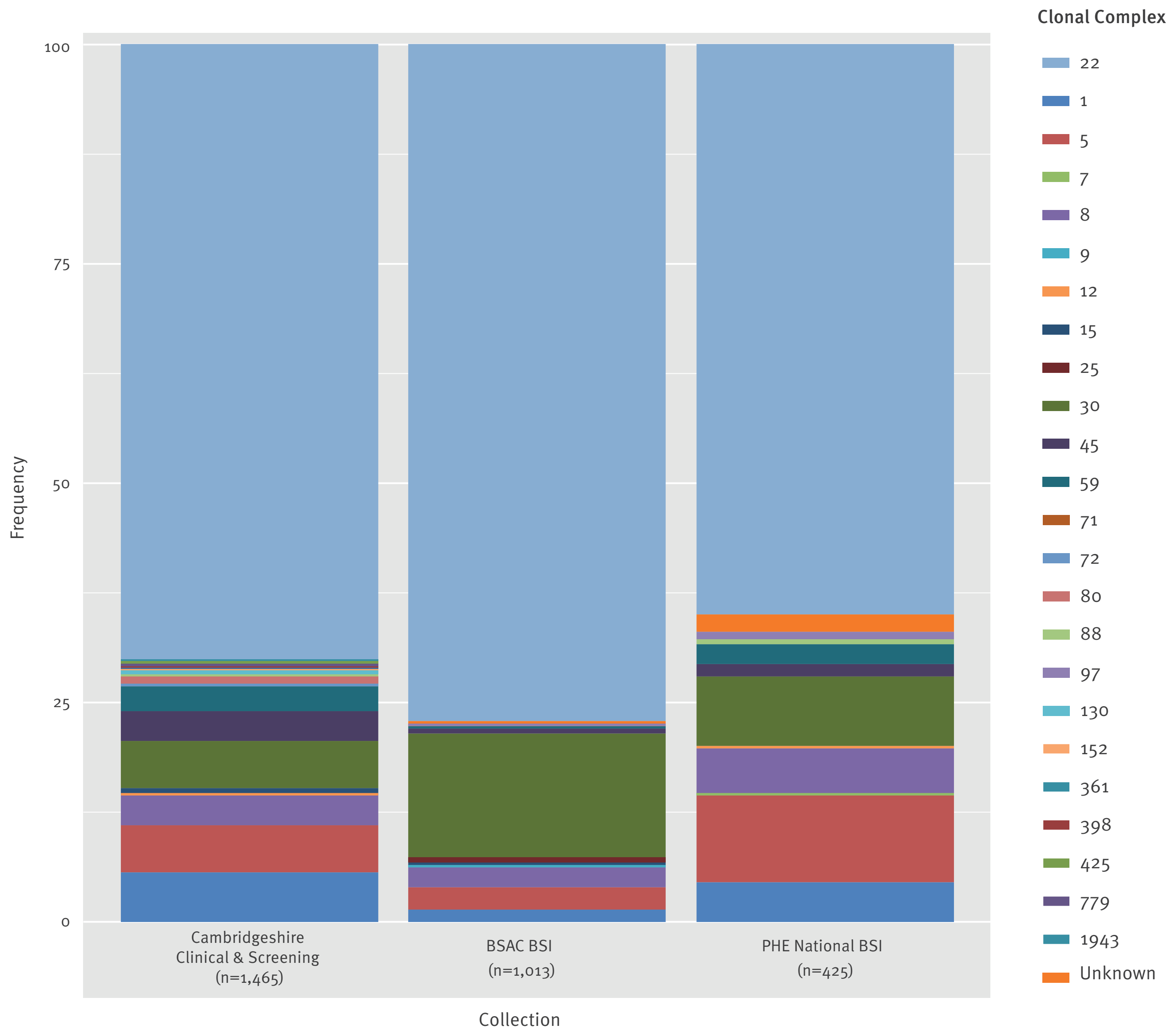
Mid-point rooted maximum likelihood tree based on SNPs in the core genome alignment generated after mapping against the reference genome EMRSA-15. The inner ring denotes the collection and the outer represents referral network of submitting laboratory. The arc indicates a large expansion in the Cambridgeshire region, which is underrepresented in the BSI only surveillance.

Isolates were *spa* genotyped using *in-silico* PCR to extract the *spa* gene X region from assembled genomes using previously described primers [23]. The *spa*-type was determined using an online *spa*-typer tool (http://spatyper.fortinbras.us/). The types generated through *spa*-genotype and laboratory determined *spa*-typing methods were compared to determine concordance.

Bacterial DNA sequences were deposited in the European Nucleotide Archive (ENA), (https://www.ebi.ac.uk/ena), under study number ERP005128. Accession numbers, details of reads, depth of coverage and N50 are provided in Supplementary Table S1. For subsequent analyses we sourced MRSA sequence data from previously published studies. These included: (i) a prospective observational cohort study of all MRSA carriage and clinical isolates submitted and processed in Cambridge University Hospital NHS Foundation Trust, Cambridgeshire, UK between 2012 and 2013 [24], (ii) MRSA bloodstream isolates collected by the BSAC BSI Surveillance Programme between 2001 and 2010 [13], (iii) USA300 isolates collected in New York, United States (US) between 2009 and 2011 [25], (iv) MRSA

Figure 3

Diversity of lineages (CC) within three isolate collections: carriage and clinical samples from the Cambridgeshire study of MRSA[a]; the national PHE BSI collection[b] and a national BSAC BSI collection[c]



BSAC: British Society of Antimicrobial Chemotherapy; BSI: blood stream infections; CC: clonal complex; MRSA: meticillin-resistant *Staphylococcus aureus*; PHE: Public Health England.

[a] Carriage and clinical samples from a 1-year-long Cambridgeshire study of MRSA (April 2012–April 2013) [24].

[b] The national PHE BSI collection (October 2012–October 2013).

[c] National BSAC BSI collection (2000–2010) [13].

isolates from outbreak investigations at a UK hospital [9,10,12].

Results

A total of 903 MRSA BSI cases were reported to MESS during the study period (Supplementary Figure S1). Gender was recorded for 98% of cases and 584 (65%) of cases were male. Age was recorded for all but two cases, with a median age of 72 years (range 0–103 years; interquartile range (IQR) 56–84 years). A total of 111 laboratories participated in the study.

A total of 559 MRSA bloodstream isolates were received. Following quality control procedures 134 isolates were excluded, and 425 isolates were included in the analysis. The reasons of exclusion were as follows: duplicate isolates (n = 50); not MRSA (n = 15); inadequate isolate growth (n = 2); isolates collected outside of the

study dates (n = 16); isolates submitted in error (n = 3); non-bloodstream isolates (n = 2); isolates from Wales (n = 28); and isolates from Northern Ireland (n = 18).

Of 903 reported BSI cases occurring in England during the study period, 47% (n = 425) had isolates that were sequenced and analysed (Figure 1). All of the 425 sequenced isolates were *mecA* positive by laboratory-PCR. PCR testing identified 8.7% (n = 37) of the isolates as PVL-positive. Based on sequence data, 65% (n = 276) were assigned to CC22. Other CCs were represented at lower frequencies: CC5 n = 42; CC30 n = 33; CC8 n = 22; CC1 n = 19; CC59 n = 9; CC45 n = 7; other/unknown CCs n = 17. The number of isolates and variation in the CCs isolated from each region is shown in Figure 1. No associations were found between particular CCs and community vs hospital onset (Supplementary Table S2).

Comparison of blood stream infection surveillance and universal methicillin-resistant *Staphylococcus aureus* sampling

We compared the most common clone in our collection, CC22 (n = 276), with CC22 genomes generated by a prospective study that sequenced MRSA isolates from every positive case (carriage and clinical samples) identified at a single diagnostic microbiology laboratory that processed samples from three hospitals and 75 general practitioner (GP) surgeries in Cambridgeshire between April 2012 and April 2013 [24]. This Cambridgeshire collection was used to represent the diversity of carriage and clinical isolates within a defined geographical area, as a national collection of carriage and clinical isolates was not available. A phylogeny was constructed for the genomes from the national BSI collection within this Cambridgeshire collection (Figure 2), in order to determine whether those isolates causing BSI were clonally related, or distributed throughout the phylogenetic tree.

As shown in Figure 2, isolates from our national MRSA BSI collection were dispersed throughout the Cambridgeshire phylogeny, ruling out any association between a particular lineage and BSI. Comparing the national BSI collection to WGS of universal sampling in Cambridgeshire also demonstrates that some lineages are under-represented when undertaking BSI-based (rather than clinical/carriage based) surveillance. For example, a large expansion (indicated with an arc on the figure) was seen in the Cambridgeshire phylogeny, with only eight of the Cambridgeshire isolates within the national MRSA BSI collection from the East of England.

To explore the effect of different sampling strategies on MRSA lineage diversity we conducted a comparison of CCs within three different MRSA collections: this national MRSA BSI collection (October 2012–October 2013), isolates from the Cambridgeshire study (April 2012–April 2013) [24], and MRSA BSI isolates from the British Society of Antimicrobial Chemotherapy (BSAC) BSI Surveillance Programme from 2000–2010 [13] (Figure 3). Despite the different sampling strategies and time frames, we found that CC22 was the dominant lineage in all collections. Both of the BSI-based collections showed a lower diversity of lineages than seen in the 1-year Cambridgeshire study. Furthermore, the BSAC collection, which collected BSI from up to 40 laboratories in the UK between 2001 and 2010, showed the most limited diversity. This may have resulted from a decline in certain lineages e.g. EMRSA-16 (CC30) during the 10-year collection period.

Contextualisation of previously recognised outbreaks

Reuter *et al.* have previously demonstrated that it is possible to use sequence data from BSAC MRSA BSI collection (2001–2010) to provide genomic context for local MRSA outbreaks within a single hospital setting [13]. We conducted a similar analysis, using the national MRSA BSI collection as context, to see if this might be feasible using a smaller sample of BSI collected during the study period of 1-year. We found that previous outbreaks in a neonatal intensive care unit [10] and a paediatric intensive care unit [9] were easily identifiable as discrete clusters, as shown in Figure 4. Furthermore, MRSA isolates from a suspected outbreak on a hepatology ward [12] were scattered throughout the phylogeny, refuting the outbreak as had been shown previously.

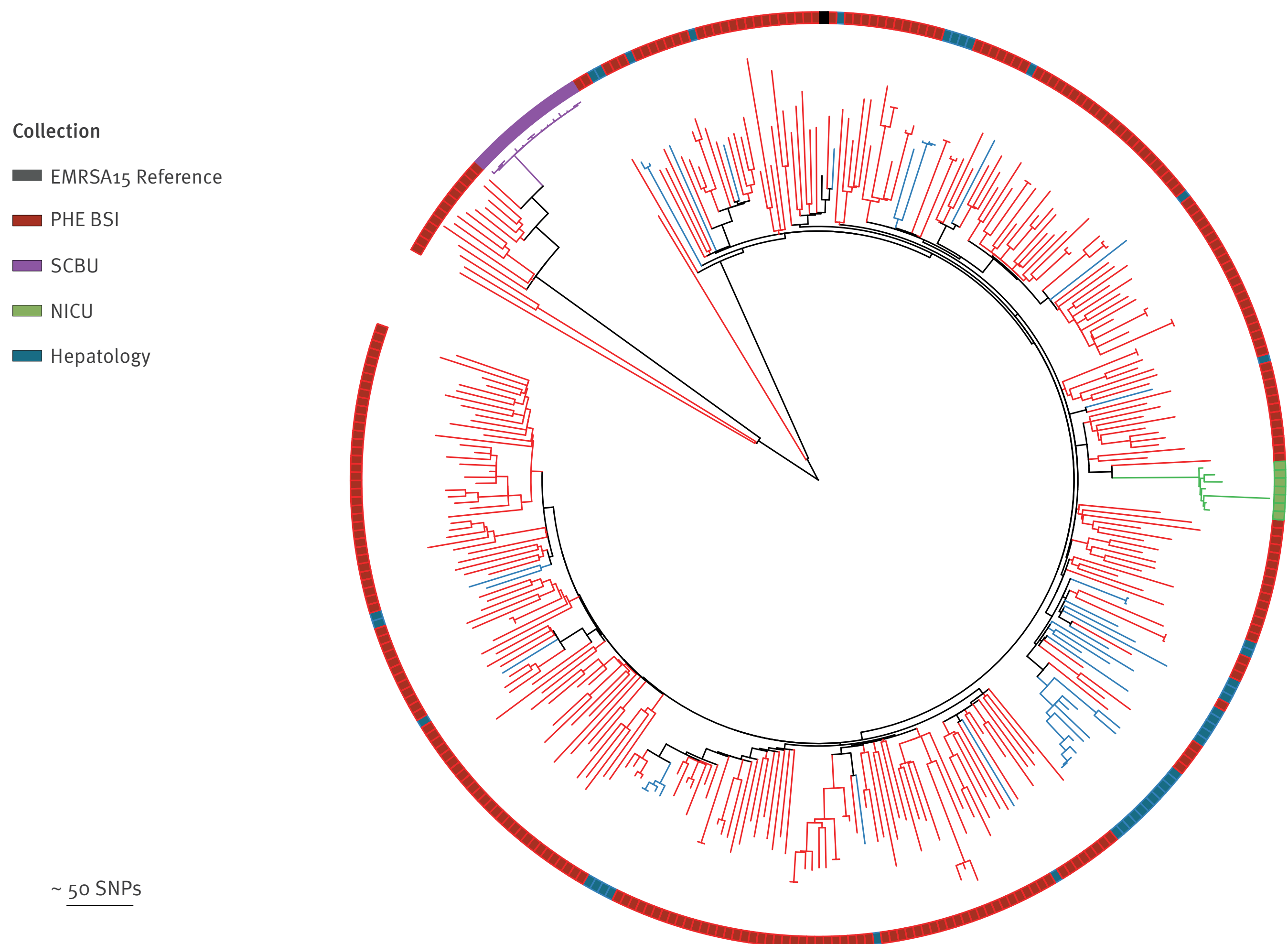
Monitoring and detection of emerging or high-risk lineages

One key aim of a national MRSA surveillance is the identification and monitoring of emerging and/or high-risk MRSA lineages. One such lineage is the USA300 lineage, which was first identified in 1999 and has subsequently caused an epidemic of skin and soft tissue infection (SSTI) in the US [26,27]. The widespread dissemination of USA300 in otherwise healthy people and its spread into hospitals has made this a high-risk strain. However, despite multiple introductions into a number of countries, genomic surveillance has shown that to date, minimal transmission of USA300 has occurred in Europe [28-32]. We examined the national MRSA BSI collection and found that six of the 22 CC8 isolates were phylogenetically defined as USA300 and were widely dispersed throughout the collection, indicating multiple introductions of USA300 into England (Figure 5). Given the observation that USA300 is commonly associated with SSTI (which are rarely sampled), and the limitations of BSI-based sampling, it is likely that the prevalence of USA300 in the UK may be higher than detected in this study.

Another potential benefit of having access to national surveillance data is the ability to identify and explore changes in molecular epidemiology on a local scale. By way of an example, we found an expansion of CC5 in South West England (Figure 6), which was genetically distinct from a CC5 expansion in Wales identified in the BSAC collection [13], despite their close geographic proximity.

Figure 4

Phylogenetic tree showing CC 22 isolates from the PHE BSI collection, England, October 2012–September 2013 (n = 276) providing contextualisation of previously published outbreaks at Cambridge University Hospitals, England (n = 64)



BSI: bloodstream infections; NICU: neonatal intensive care unit; PHE: Public Health England; SCBU: special care baby unit; SNPs: single nucleotide polymorphisms.

Mid-point rooted maximum likelihood tree based on SNPs in the core genome alignment generated after mapping against the reference genome EMRSA-15. Isolates from NICU (n = 7) are indicated in green [10]. Isolates from an outbreak on a SCBU which extended into epidemiologically linked cases in the community (n = 15) are indicated in purple [9]. Isolates from a suspected but disproven outbreak on a hepatology ward (n = 42) are indicated in blue [12].

Backward compatibility of typing methods

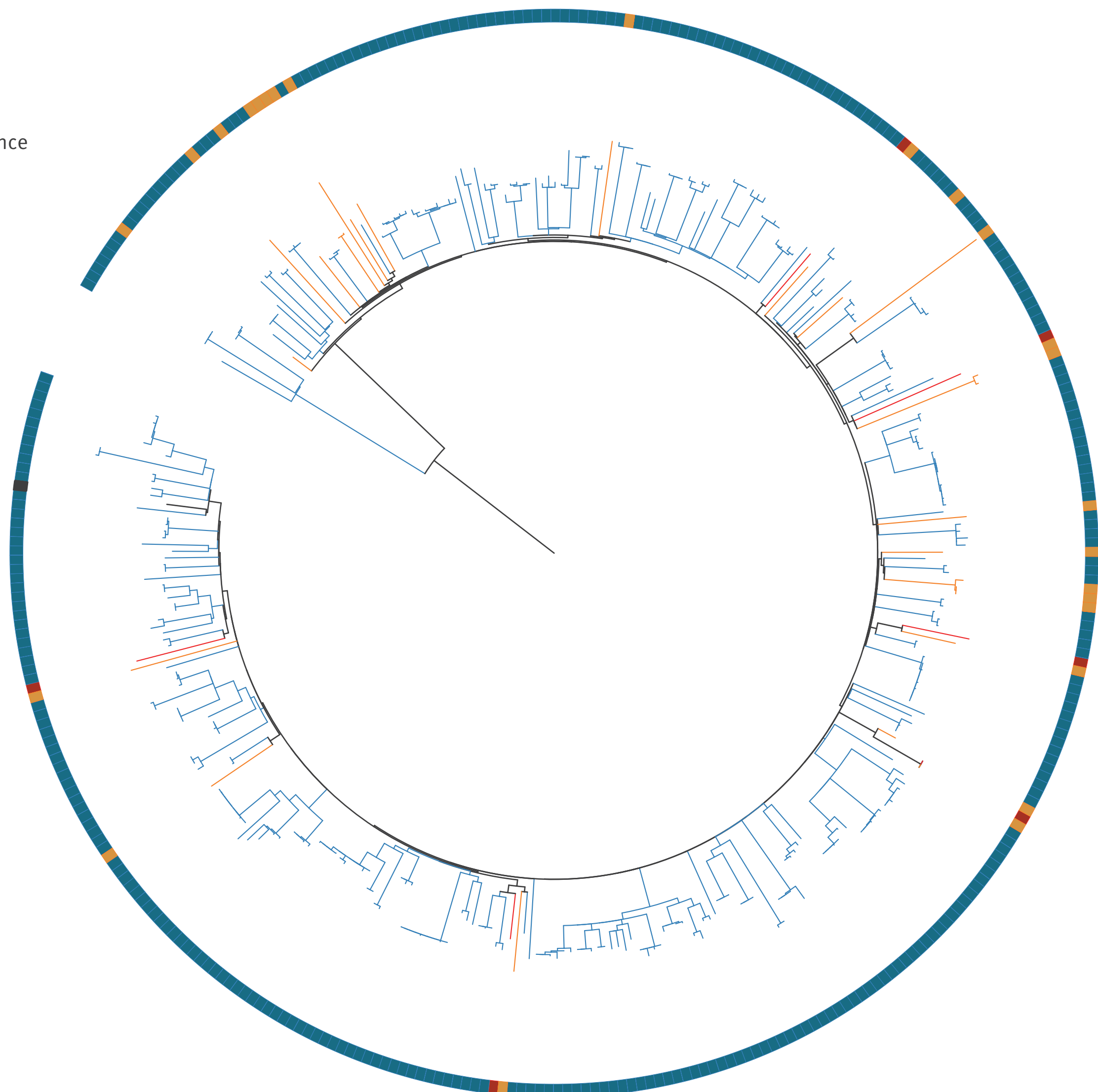
Globally, PCR determination of *spa*-type is a commonly used typing method. However, as some laboratories transition to WGS-based typing, it is important that typing methods remain compatible. We examined the concordance between these two methods in the national MRSA BSI collection. Of the 425 isolates we found a 98.4% concordance rate (Supplementary Table S2), comparable to previous studies [33,34]. Of the seven isolates with discordant results, there were deletions/rearrangements within the *spa* gene of the short-read assemblies that resulted in loss of sequence complementary to forward primer, and thus failure to *in silico* amplify the gene region targeted by genomic *spa*-typing.

Discussion

Mandatory enhanced surveillance for MRSA BSI in England has provided in-depth information on the national decline of MRSA BSI and the changes in patient-level epidemiology that have accompanied it [2]. However, without characterisation of systematically collected isolates, bacterial molecular epidemiology cannot be studied. This study aimed to investigate whether it was feasible to undertake combined epidemiological and genomic surveillance of MRSA bloodstream infections in England in order to address this issue.

Figure 5

Phylogenetic tree showing USA300 isolates from the PHE BSI collection, alongside previously published USA300 isolates from a universal sample collection in Cambridgeshire and from the United States



PHE: Public Health England; SNPs: single nucleotide polymorphisms; USA: United States of America.

PHE BSI USA300 isolates (n = 6) are indicated in red. USA300 isolates from a universal sample collection in Cambridgeshire (n = 24) are indicated in orange [24] and from the US (n = 348) in blue [25].

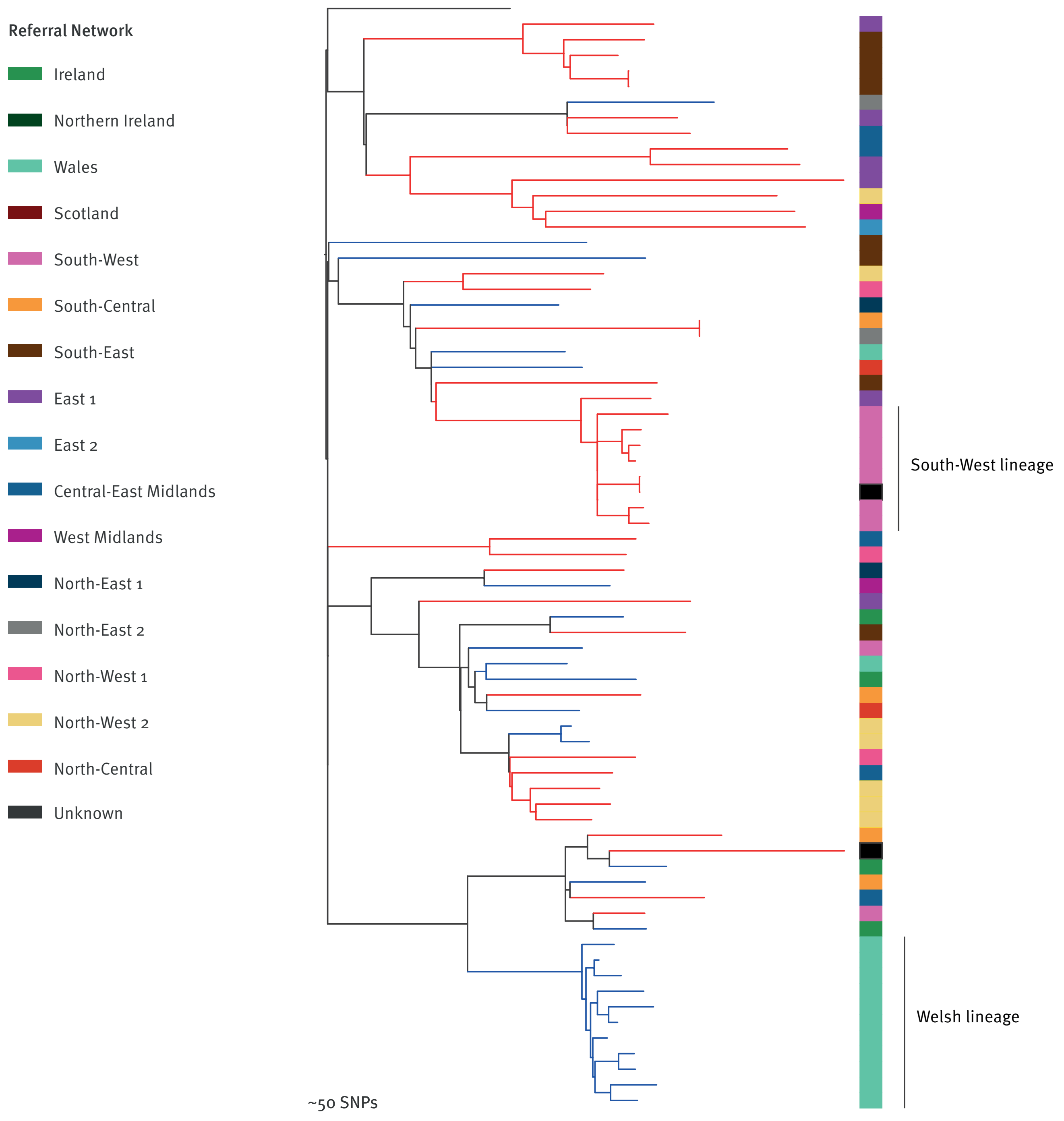
Mid-point rooted maximum likelihood tree based on SNPs in the core genome alignment generated after mapping against the reference genome FPR3757 (black).

We encountered some challenges including obtaining bloodstream isolates from all participating hospitals (as submission was voluntary) and integrating two datasets collected through different methods (epidemiological data collected through an online database submission and isolates sent with written information via post/courier). Despite this, we demonstrated the feasibility of this approach. We were able to construct the known population structure and diversity of MRSA in England, even with an incomplete collection of bloodstream isolates collected over a 1-year period. We found a greater diversity of clones than that seen in a 10-year national collection of MRSA bloodstream isolates (BSAC collection) with a limited sampling strategy, but less diversity than that seen in a 1-year regional collection of carriage and clinical isolates (Cambridgeshire study [24]). A sensible first step in MRSA surveillance is to assess existing genomic diversity [16] and our study demonstrates that this can be achieved and could feasibly be extended over time to generate a comprehensive national genomic database to monitor changes in clonal diversity.

Prior to April 2017, all MRSA BSI isolates submitted to PHE were routinely characterised by *spa*-typing and PCR to confirm species identification alongside

Figure 6

Phylogenetic tree showing PHE CC5 bloodstream infection isolates, 1 October 2012–30 September 2013 (n=42) and CC5 isolates from the previously published BSAC collection of bloodstream infections, 2000–2010 (n=28)



BSAC: British Society of Antimicrobial Chemotherapy; BSI: blood stream infections; CC: clonal complex; PHE: Public Health England; SNPs: single nucleotide polymorphisms.

PHE BSI CC5 isolates (n = 42) indicated by red branches. National BSI BSAC collection (n = 28) indicated by blue branches [13].

Mid-point rooted maximum likelihood tree based on SNPs in the core genome alignment generated after mapping against the CC5 reference genome N315. Coloured bar represents referral network of submitting laboratory. Expansions within the South West England and Welsh regions are highlighted.

determination of *mecA/C* and *luk-PV* status [17,18]. As typing methods evolve and WGS becomes increasingly routine, backward compatibility with previous methods ensures the continued utility of typed historical collections. Laboratory *spa*-typing and *spa*-genotyping from short read WGS data have been shown to be largely comparable in a limited number of studies [33,34], despite the high density of repeats within the *spa* gene region. We showed over 98% concordance between laboratory and genomic *spa*-typing methods which, reassuringly, confirms compatibility with historical data.

A further potential benefit of prospective sequencing of MRSA bloodstream isolates and a centralised national database is the ability to provide genomic context to confirm or refute outbreaks on a local or a national scale. This would be an invaluable resource as long as there is open access to anonymised (non-identifiable) data and to bioinformatics tools to analyse them rapidly and easily. The challenges will be to ensure standardised methods, development of strategies to avoid duplication of samples and establishment of a large, comprehensive, open access, anonymised database where data could be deposited, curated and accessible for public health benefit. Web-based, open-access software packages that are potentially suitable for this purpose are already being developed [35,36]. Apart from the ability to detect emerging or potentially high-risk MRSA clones retrospectively, on-going sampling and analysis will enable detection in real-time.

In this study, we found that that the high-risk USA300 lineage, an epidemic cause of SSTI in the US [26], has spread to the UK and is causing bloodstream infections across England. While the genomic data suggest multiple introductions of USA300, the use of BSI rather than clinical isolate-based surveillance limits our ability to analyse this further. However, using the PHE BSI collection it was possible to identify a local expansion of CC5 causing BSI in South West England, where local investigations suggest this clone has been causing excess disease [37]. Thus, timely, routine WGS of PHE BSI isolates combined with local epidemiological data could potentially identify novel and/or pathogenic lineages in real time and could be used to trigger local /regional investigations and interventions.

A major advantage of sequencing MRSA isolates is the ability to share and collate genome sequence data to build up national and international databases. A number of BSI surveillance systems already exist e.g. the English mandatory enhanced surveillance system, the voluntary British Society of Antimicrobial Chemotherapy BSI Surveillance Programme and the voluntary European Antimicrobial Resistance Surveillance Network. While each system has different aims and objectives, sampling criteria and data collection methods, the digital interchangeability of sequence data creates an opportunity to collaborate and share genome sequence data while producing a sustainable, on-going resource if the isolates were sequenced. The challenges will be to ensure standardised methods, development of strategies to avoid duplication of samples and establishment of a large, comprehensive, open-access database where anonymised data could be deposited, curated, and accessible for public health benefit.

We acknowledge several limitations in our study. The systems for collecting epidemiological data and bacterial isolates were separate and different, leading to high rates of sample exclusion. This challenge of capturing and integrating both types of data could be overcome in practice by submitting epidemiological and laboratory data to a single data collection system. Submission of bloodstream isolates was voluntary, with many reported cases having no corresponding isolate referred for characterisation; this may have introduced bias into the analysis. This could be addressed by having mandatory submission of isolates for all reported cases. Finally, we did not conduct a cost/benefit analysis of this approach. Despite these limitations, we have demonstrated that prospective epidemiological and genomic surveillance of MRSA bloodstream infections is feasible, has numerous potential benefits and could provide a valuable public health resource in England and beyond.

Acknowledgements

We thank participating NHS hospital and laboratory staff for submitting epidemiological data and bacterial isolates to the mandatory enhanced surveillance system and the National Staphylococcal Reference Laboratory, respectively. We also thank the Sequencing and Pathogen Informatics groups at the Wellcome Sanger Institute for their support.

Funding: This work was supported by grants from the UK Clinical Research Collaboration Translational Infection Research Initiative, and the Medical Research Council (Grant Number G1000803) with contributions to the Grant from the Biotechnology and Biological Sciences Research Council, the National Institute for Health Research on behalf of the Department of Health, and the Chief Scientist Office of the Scottish Government Health Directorate (to Prof. Peacock); and the Wellcome Trust (to Prof. Parkhill [Grant 098051], Prof. Peacock). MST is a Wellcome Trust Clinical PhD Fellow at the University of Cambridge. MET is a Clinician Scientist Fellow, supported by the Academy of Medical Sciences and the Health Foundation, and by the National Institute for Health Research Cambridge Biomedical Research Centre. FC is supported by the Wellcome Trust (201344/Z/16/Z).

Conflict of interest

JP, SJP and FC are paid consultants to Next Gen Diagnostics LLC. All other authors declare no conflicts of interest.

Authors’ contributions

MET designed the study, wrote the protocol and obtained ethical and R&D approvals for the study. MET, JP and SJP supervised the study. BB and HJW performed the laboratory work in Cambridge. RH and AK co-ordinated the collection and characterisation of isolates and provided the

epidemiological and laboratory data from PHE Colindale. MST, MET, SR, FC, EH and DJ provided data, advice or performed data analyses. MST and MET wrote the first draft of the manuscript; all authors contributed to and approved the final manuscript.

References

1. Duerden B, Fry C, Johnson AP, Wilcox MH. The Control of Methicillin-Resistant Staphylococcus aureus Blood Stream Infections in England. Open Forum Infect Dis. 2015;2(2):ofv035. https://doi.org/10.1093/ofid/ofv035 PMID: 26380336
2. Johnson AP, Davies J, Guy R, Abernethy J, Sheridan E, Pearson A, et al. Mandatory surveillance of methicillin-resistant Staphylococcus aureus (MRSA) bacteraemia in England: the first 10 years. J Antimicrob Chemother. 2012;67(4):802-9. https://doi.org/10.1093/jac/dkr561 PMID: 22223229
3. Coia JE, Duckworth GJ, Edwards DI, Farrington M, Fry C, Humphreys H, et al. Joint Working Party of the British Society of Antimicrobial ChemotherapyHospital Infection SocietyInfection Control Nurses Association. Guidelines for the control and prevention of meticillin-resistant Staphylococcus aureus (MRSA) in healthcare facilities. J Hosp Infect. 2006;63(Suppl 1):S1-44. https://doi.org/10.1016/j.jhin.2006.01.001 PMID: 16581155
4. Wilson M. Addressing the problems of long-term urethral catheterization: Part 1. Br J Nurs. 2011;20(22):1418-20. https://doi.org/10.12968/bjon.2011.20.22.1418 PMID: 22241489
5. Wyllie D, Paul J, Crook D. Waves of trouble: MRSA strain dynamics and assessment of the impact of infection control. J Antimicrob Chemother. 2011;66(12):2685-8. https://doi.org/10.1093/jac/dkr392 PMID: 21948966
6. Lawes T, López-Lozano JM, Nebot C, Macartney G, Subbarao-Sharma R, Dare CR, et al. Turning the tide or riding the waves? Impacts of antibiotic stewardship and infection control on MRSA strain dynamics in a Scottish region over 16 years: non-linear time series analysis. BMJ Open. 2015;5(3):e006596. https://doi.org/10.1136/bmjopen-2014-006596 PMID: 25814495
7. Price JR, Didelot X, Crook DW, Llewelyn MJ, Paul J. Whole genome sequencing in the prevention and control of Staphylococcus aureus infection. J Hosp Infect. 2013;83(1):14-21. https://doi.org/10.1016/j.jhin.2012.10.003 PMID: 23164609
8. Garvey MI, Pichon B, Bradley CW, Moiemen NS, Oppenheim B, Kearns AM. Improved understanding of an outbreak of meticillin-resistant Staphylococcus aureus in a regional burns centre via whole-genome sequencing. J Hosp Infect. 2016;94(4):401-4. https://doi.org/10.1016/j.jhin.2016.09.013 PMID: 27729168
9. Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis. 2013;13(2):130-6. https://doi.org/10.1016/S1473-3099(12)70268-2 PMID: 23158674
10. 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. https://doi.org/10.1056/NEJMoa1109910 PMID: 22693998
11. Price JR, Golubchik T, Cole K, Wilson DJ, Crook DW, Thwaites GE, et al. Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit. Clin Infect Dis. 2014;58(5):609-18. https://doi.org/10.1093/cid/cit807 PMID: 24336829
12. Török ME, Harris SR, Cartwright EJ, Raven KE, Brown NM, Allison ME, et al. Zero tolerance for healthcare-associated MRSA bacteraemia: is it realistic? J Antimicrob Chemother. 2014;69(8):2238-45. https://doi.org/10.1093/jac/dku128 PMID: 24788657
13. Reuter S, Török ME, Holden MTG, Reynolds R, Raven KE, Blane B, et al. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. Genome Res. 2016;26(2):263-70. https://doi.org/10.1101/gr.196709.115 PMID: 26672018
14. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, et al. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. Genome Res. 2013;23(4):653-64. https://doi.org/10.1101/gr.147710.112 PMID: 23299977
15. Grundmann H, Schouls LM, Aanensen DM, Pluister GN, Tami A, Chlebowicz M, et al. ESCMID Study Group on Molecular Epidemiological MarkersEuropean Staphylococcal Reference Laboratory Working Group. The dynamic changes of dominant clones of Staphylococcus aureus causing bloodstream infections in the European region: results of a second structured survey. Euro Surveill. 2014;19(49):20987. https://doi.org/10.2807/1560-7917.ES2014.19.49.20987 PMID: 25523972
16. Planet PJ. Life After USA300: The Rise and Fall of a Superbug. J Infect Dis. 2017;215(suppl_1):S71-S77.
17. Holmes A, Ganner M, McGuane S, Pitt TL, Cookson BD, Kearns AM. Staphylococcus aureus isolates carrying Panton-Valentine leucocidin genes in England and Wales: frequency, characterization, and association with clinical disease. J Clin Microbiol. 2005;43(5):2384-90. https://doi.org/10.1128/JCM.43.5.2384-2390.2005 PMID: 15872271
18. Pichon B, Hill R, Laurent F, Larsen AR, Skov RL, Holmes M, et al. Development of a real-time quadruplex PCR assay for simultaneous detection of nuc, Panton-Valentine leucocidin (PVL), mecA and homologue mecALGA251. J Antimicrob Chemother. 2012;67(10):2338-41. https://doi.org/10.1093/jac/dks221 PMID: 22687894
19. Reuter S, Ellington MJ, Cartwright EJ, Köser CU, Török ME, Gouliouris T, et al. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern Med. 2013;173(15):1397-404. https://doi.org/10.1001/jamainternmed.2013.7734 PMID: 23857503
20. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb Genom. 2016;2(8):e000083. https://doi.org/10.1099/mgen.0.000083 PMID: 28348874
21. Page AJ, Taylor B, Keane JA. Multilocus sequence typing by blast from de novo assemblies against PubMLST. The journal of Open Source Software. 2016;8(1):118. https://doi.org/10.21105/joss.00118
22. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005;21(4):456-63. https://doi.org/10.1093/bioinformatics/bti191 PMID: 15608047
23. Kahl BC, Mellmann A, Deiwick S, Peters G, Harmsen D. Variation of the polymorphic region X of the protein A gene during persistent airway infection of cystic fibrosis patients reflects two independent mechanisms of genetic change in Staphylococcus aureus. J Clin Microbiol. 2005;43(1):502-5. https://doi.org/10.1128/JCM.43.1.502-505.2005 PMID: 15635028
24. Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. Sci Transl Med. 2017;9(413):eaak9745. PMID: 29070701
25. Uhlemann A-C, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MTG, et al. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci USA. 2014;111(18):6738-43. https://doi.org/10.1073/pnas.1401006111 PMID: 24753569
26. David MZ, Daum RS. Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. Clin Microbiol Rev. 2010;23(3):616-87. https://doi.org/10.1128/CMR.00081-09 PMID: 20610826
27. Centers for Disease Control and Prevention (CDC). Methicillin-resistant Staphylococcus aureus skin or soft tissue infections in a state prison--Mississippi, 2000. MMWR Morb Mortal Wkly Rep. 2001;50(42):919-22. PMID: 11699844
28. Seidl K, Leimer N, Palheiros Marques M, Furrer A, Senn G, Holzmann-Bürgel A, et al. USA300 methicillin-resistant Staphylococcus aureus in Zurich, Switzerland between 2001 and 2013. Int J Med Microbiol. 2014;304(8):1118-22. https://doi.org/10.1016/j.ijmm.2014.08.005 PMID: 25200859
29. Baud O, Giron S, Aumeran C, Mouly D, Bardon G, Besson M, et al. First outbreak of community-acquired MRSA USA300 in France: failure to suppress prolonged MRSA carriage despite decontamination procedures. Eur J Clin Microbiol Infect Dis. 2014;33(10):1757-62. https://doi.org/10.1007/s10096-014-2127-6 PMID: 24816900
30. Blanco R, Tristan A, Ezpeleta G, Larsen AR, Bes M, Etienne J, et al. Molecular epidemiology of Panton-Valentine leukocidin-positive Staphylococcus aureus in Spain: emergence of the USA300 clone in an autochthonous population. J Clin Microbiol. 2011;49(1):433-6. https://doi.org/10.1128/JCM.02201-10 PMID: 21068288
31. van der Mee-Marquet N, Poisson DM, Lavigne JP, Francia T, Tristan A, Vandenesch F, et al. The incidence of Staphylococcus aureus ST8-USA300 among French pediatric inpatients is rising. Eur J Clin Microbiol Infect Dis. 2015;34(5):935-42. https://doi.org/10.1007/s10096-014-2308-3 PMID: 25575950

32. Toleman MS, Reuter S, Coll F, Harrison EM, Blane B, Brown NM, et al. Systematic Surveillance Detects Multiple Silent Introductions and Household Transmission of Methicillin-Resistant Staphylococcus aureus USA300 in the East of England. J Infect Dis. 2016;214(3):447-53. https://doi.org/10.1093/infdis/jiw166 PMID: 27122590
33. Bartels MD, Petersen A, Worning P, Nielsen JB, Larner-Svensson H, Johansen HK, et al. Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant Staphylococcus aureus. J Clin Microbiol. 2014;52(12):4305-8. https://doi.org/10.1128/JCM.01979-14 PMID: 25297335
34. Bletz S, Mellmann A, Rothgänger J, Harmsen D. Ensuring backwards compatibility: traditional genotyping efforts in the era of whole genome sequencing. Clin Microbiol Infect. 2015;21(4):347.e1-4. https://doi.org/10.1016/j.cmi.2014.11.005 PMID: 25658529
35. Argimón S, Abudahab K, Goater RJ, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2(11):e000093. https://doi.org/10.1099/mgen.0.000093 PMID: 28348833
36. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11(1):595. https://doi.org/10.1186/1471-2105-11-595 PMID: 21143983
37. Packer STS, Buunaaisie C, Neale J, Hopewell-Kelly N, Telfer M, Williams OM, et al. A prevalence survey investigating factors associated with MRSA colonisation in people who inject drugs in Bristol, UK. Poster presentation at Research and Applied Epidemiology Scientific Conference; 2017 Mar 21 Warwick University, Coventry, UK.

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Research

Whole genome sequencing of *Salmonella* Chester reveals geographically distinct clusters, Norway, 2000 to 2016

Lotta Siira[1,2], Umaer Naseer[1], Kristian Alfsnes[1], Nils Olav Hermansen[1,3], Heidi Lange[1], Lin T Brandal[1]
1. Division for Infection Control and Environmental Health, Norwegian Institute of Public Health (NIPH), Oslo, Norway
2. European Program for Public Health Microbiology Training (EUPHEM), European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
3. Department of Microbiology, Oslo University Hospital, Oslo, Norway

Correspondence: Lotta Siira (lols@fhi.no)

Citation style for this article:
Siira Lotta, Naseer Umaer, Alfsnes Kristian, Hermansen Nils Olav, Lange Heidi, Brandal Lin T. Whole genome sequencing of Salmonella Chester reveals geographically distinct clusters, Norway, 2000 to 2016. Euro Surveill. 2019;24(4):pii=1800186. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800186

Article submitted on 13 Apr 2018 / accepted on 06 Nov 2018 / published on 24 Jan 2019

Introduction: During summer 2016, Norway observed an increase in *Salmonella enterica* subsp. *enterica* serovar Chester cases among travellers to Greece. Aim: Our aim was to investigate genetic relatedness of *S.* Chester for surveillance and outbreak detection by core genome multilocus sequence typing (cgMLST) and compare the results to genome mapping. **Methods:** We included *S.* Chester isolates from 51 cases of salmonellosis between 2000 and 2016. Paired-end sequencing (2 × 250 bp) was performed on Illumina MiSeq. Genetic relatedness by cgMLST for *Salmonellaenterica* subsp. *enterica,* including 3,002 genes and seven housekeeping genes, was compared by reference genome mapping with CSI Phylogeny version 1.4 and conventional MLST. **Results:** Confirmed travel history was available for 80% of included cases, to Europe (n = 13), Asia (n = 12) and Africa (n = 16). Isolates were distributed into four phylogenetic clusters corresponding to geographical regions. Sequence type (ST) ST411 and a single-locus variant ST5260 (n = 17) were primarily acquired in southern Europe, ST1954 (n = 15) in Africa, ST343 (n = 11) and ST2063 (n = 8) primarily in Asia. Part of the European cluster was further divided into a Greek (n = 10) and a Cypriot (n = 4) cluster. All isolates in the African cluster displayed resistance to ≥ 1 class of antimicrobials, while resistance was rare in the other clusters. **Conclusion:** Whole genome sequencing of *S.* Chester in Norway showed four geographically distinct clusters, with a possible outbreak occurring during summer 2016 related to Greece. We recommend public health institutes to implement cgMLST-based real-time *Salmonellaenterica* surveillance for early and accurate detection of future outbreaks and further development of cluster cut-offs.

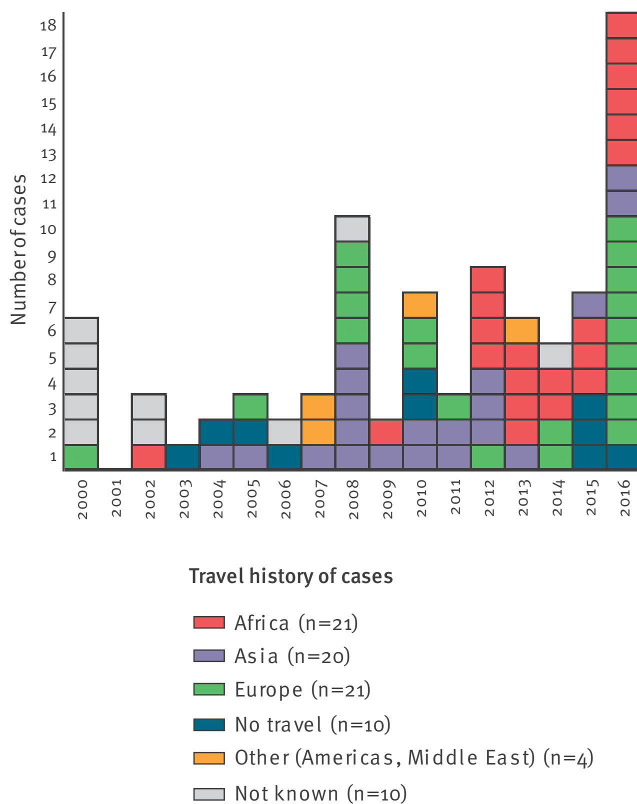
Introduction

Salmonellosis is characterised by gastroenteritis with acute onset of fever, abdominal pain, diarrhoea, nausea and occasionally vomiting and is one of the most commonly reported food-borne diseases in Europe. In 2016, 20 confirmed salmonellosis cases per 100,000 population were reported in the European Union (EU) [1]. During the summers of 2014 and 2015, several European countries reported an increase in cases with salmonellosis caused by *Salmonella enterica* subsp. *enterica* serovar Chester. This multi-country outbreak was associated with travel to Morocco and was probably linked to multiple food sources [2]. Previously, human cases of salmonellosis from *S.*Chester had rarely been reported, but after this outbreak, *S.* Chester was included among the 20 most common *Salmonella* serovars causing infections in humans in Europe in 2014 [3]. *S.* Chester has since accounted for 0.4% of the annually reported salmonellosis cases in Europe [1]. Outbreaks caused by *S.* Chester have also been reported elsewhere. In 2010, *S.* Chester was implicated in two outbreaks in North America: in Canada, head cheese (brawn) was identified as the source [4], and a multi-state outbreak in the United States (US) was associated with frozen meals [5]. In China, *S.* Chester was isolated in a multi-serovar *Salmonella* outbreak in 2012, where egg sandwiches were implicated as the main vehicle [6]. In Australia, *S.* Chester outbreaks associated with turtle meat and municipal water were described in 1998 and 2005, respectively [7,8].

Salmonellosis has been notifiable to the Norwegian Surveillance System for Communicable Diseases (MSIS, http://www.msis.no/) since 1977, and the corresponding isolates are sent to the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health (NIPH). During the period from 2000 to 2016, the number of *S.* Chester cases by year in Norway has ranged between zero and 18. A travel history was confirmed in 78% of the cases; Europe, Asia and Africa were approximately equally represented as travel destinations of the cases (Figure 1). During summer 2016, we observed an increase in cases with a history of travel to the island of Rhodes, Greece.

Figure 1

Salmonella Chester cases, by travel history and year, Norway, 2000–2016 (n = 86)



This information was shared with other European countries through the Epidemic Intelligence Information System (EPIS) of the European Centre for Disease Prevention and Control (ECDC), but the communication returned no reports of cases outside Norway.

In this study, we studied the genomic relatedness of *Salmonella* Chester isolates by whole genome sequencing (WGS) analysed by a core genome multilocus sequence typing (cgMLST) scheme and compared with results obtained by single nucleotide polymorphism (SNP)-based reference genome mapping. Our aim was to identify if the cases with a history of travel to Greece were part of the Moroccan outbreak cluster and to examine the molecular epidemiology of isolates with different geographical origins. Comparisons of cgMLST- and SNP-based results are important as the use of WGS is increasingly used in public health. This requires information on the diversity of sequences within a species, serovar and previously defined genotypes, to determine appropriate cut-offs for clusters and outbreaks. Our study contributes to this body of knowledge. As Norwegians travel frequently both within Europe and outside the continent [9] and the majority of the *Salmonella* cases identified in Norway report a history of travel, our surveillance data are well positioned to give insight into the internationally circulating *S.* Chester strains.

Methods

Cases and isolates

Fifty-one of 86 non-duplicate *S.* Chester isolates from the national strain collection at the National Reference Laboratory for Enteropathogenic Bacteria at NIPH from the years 2000 to 2016 were included in the study. These consisted of all isolates from the years 2014 to 2016 (n = 30) and a selection of older isolates from the period 2000 to 2013 (n = 21) chosen so that they were representative of the travel history of all cases. Isolates from all cases reporting travel to Greece, Cyprus or Morocco in the years 2000 to 2016 were included.

Epidemiological investigations

S. Chester cases with a history of travel to Greece during summer 2016 were interviewed to obtain more detailed information on travel destination, dates of stay, accommodation, travel agency and foods consumed at the destination.

Serotyping

Serotypes were confirmed by agglutination tests with antisera (Sifin Diagnostics GmbH, Berlin, Germany and SSI, Statens Serum Institut, Hillerød, Denmark) according to the White–Kauffmann scheme [10]. The SeqSero online tool, version 1.0 (http://denglab.info/SeqSero) was used to identify the serotype from the raw sequence reads [11].

Whole genome sequencing

DNA extraction was performed by MagNAPure 96 (Roche Molecular Systems Inc., Pleasanton, US). KAPA HyperPlus (Kapa Biosystems, Wilmington, US) was used for library preparation and Agencourt AMPure XP (Beckmann Coulter Life Sciences, Indianapolis, US) for removal of adaptor dimers. WGS was performed as paired-end (250 bp × 2) sequencing on the MiSeq (Illumina, Inc., San Diego, US) platform aiming for coverage of › 50×. Quality control of the raw reads was done through FastQC. The sequences were submitted to the European Nucleotide Archive (ENA) under the access number PRJEB30485.

Multilocus sequence typing

Genotyping by the *Salmonella enterica* seven-gene MLST scheme was performed through EnteroBase in the SeqSphere+ software, version 4.0 (Ridom GmbH, Münster, Germany) based on the Achtman scheme.

Core genome multilocus sequence typing

All analyses were performed using SeqSphere+. Briefly, raw sequence reads were trimmed until an average Phred base quality of ≥ 30 was reached in a window of 20 bases, and de novo assembly was performed using Velvet version 1.1.04 with default settings. We used the SeqSphere+ integrated cgMLST scheme developed by Alikhan et al. for EnteroBase (https://enterobase.warwick.ac.uk/) [12], with allele calling procedure with a minimum accepted BLAST identity of 80%, no BLASTp search, frame-shift detection turned on and

Table A

Characteristics of *Salmonella* Chester strains included in the study, Norway, 2000–2016 (n = 51)

WGS cluster[a]	ST	Travel history	Year	Resistance determinants identified through ResFinder	PlasmidFinder and pMLST result	Phenotypic antimicrobial susceptibility				
						BL	CHL	QNL	SXT	TET
Europe (n = 17)	ST411 (n = 14)	Greece (n = 8)	2016	*QnrB19*	IncI1, IncX3(pEC14), Col(pVC)	S	ND	S	S	ND
			2016	*QnrB19*	Col(pVC)	S	ND	S	S	ND
			2016	*QnrB19*	None	S	S	S	S	S
			2016	*aph(6)-Id*	Col(pVC)	S	S	I	S	S
			2016	None	Col(pVC)	S	S	I	S	S
			2016	None	None	S	ND	S	S	ND
			2016	None	None	S	ND	S	S	ND
			2016	None	None	S	ND	S	S	ND
		None (n = 3)	2006	None	None	S	S	S	S	S
			2015	*QnrB19*	Col(pVC)	S	ND	I	S	ND
			2016	None	Col(pVC)	S	ND	S	S	ND
		Southern Europe (n = 1)	2016	None	None	S	ND	S	S	ND
		Ivory Coast (n = 1)	2012	None	None	S	S	I	S	S
		Cyprus (n = 1)	2000	None	IncI2	S	S	S	S	S
	ST5260 (n = 3)	Cyprus (n = 3)	2010	None	None	S	S	S	S	S
			2014	None	Col(pVC)	S	ND	I	S	ND
			2014	*aph(6)-Id, strA, tet(A)*	IncFII, Col(pVC)	S	ND	I	S	ND
Africa (n = 15)	ST1954 (n = 15)	Morocco (n = 14)	2013	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	Col(pVC), IncN_ST7	S	R	R	R	R
			2013	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	Col(pVC), IncN_ST7	S	R	R	R	R
			2013	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	IncN_ST7	S	R	R	R	R
			2014	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	Col(pVC), IncN_ST7	S	R	R	R	R
			2014	*aph(6)-Id, aph(3")-Ib, QnrS1, sul2, tet(A), dfrA14*	Col(pVC)	S	ND	R	R	ND
			2015	*aph(6)-Id, aph(3")-Ib, QnrB19, sul2, tet(A), dfrA14*	Col(pVC)	S	ND	R	R	ND
			2015	*aph(6)-Id, aph(3")-Ib, QnrB19, sul2, tet(A), dfrA14*	Col(pVC)	S	ND	R	R	ND
			2015	*aph(6)-Id, aph(3")-Ib, QnrB19, sul2, tet(A), dfrA14*	Col(pVC)	S	ND	R	R	ND
			2016	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	IncN_ST7	S	R	R	R	R
			2016	*aph(6)-Id, aph(3")-Ib, QnrS1, floR, sul2, tet(A), dfrA14*	IncN_ST7	S	R	R	R	R
			2016	*aph(6)-Id, aph(3")-Ib, sul2, tet(A), dfrA14*	Col(pVC)	S	S	S	R	R
			2016	*aph(6)-Id, aph(3")-Ib, sul2, tet(A), dfrA14*	Col(pVC)	S	ND	S	R	ND
			2016	*QnrB19*	Col(pVC)	S	S	R	S	S
			2016	*QnrB19*	Col(pVC)	S	S	R	S	S
		Senegal (n = 1)	2012	*aph(6)-Id, aph(3")-Ib sul2, tet(A), dfrA14*	IncI1_ST21	S	S	S	R	R

BL: β-lactam antibiotics; CHL: chloramphenicol; I: intermediate resistance; ND: not done; QNL: quinolone; R: resistant; S: susceptible; ST: sequence type; SXT: trimethoprim-sulfamethoxazole; TET: tetracycline; WGS: whole genome sequencing.

[a] WGS clusters are named for the geographical regions where the majority of cases had travelled. Asia 1 and Asia 2 refer to two separate phylogenetic clusters we identified among the cases with travel history to Asia.

Table B

Characteristics of *Salmonella* Chester strains included in the study, Norway, 2000–2016 (n = 51)

WGS cluster[a]	ST	Travel history	Year	Resistance determinants identified through ResFinder	PlasmidFinder and pMLST result	Phenotypic antimicrobial susceptibility				
						BL	CHL	QNL	SXT	TET
Asia 1[a] (n = 11)	ST343 (n = 11)	Thailand (n = 6)	2008	*bla*$_{TEM}$*-1B, QnrS1, gyrAp*.S83F	IncN	R	S	R	S	S
			2008	None	None	S	S	S	S	S
			2009	None	None	S	S	S	S	S
			2010	*gyrA* p.S83F	None	S	S	R	S	S
			2012	None	None	S	S	I	S	S
			2013	*aadA1, sul3*	IncI1_ST3	S	S	S	S	S
		None (n = 5)	2003	*bla*$_{TEM-1B}$, *QnrS1*	IncN	R	S	R	S	S
			2010	None	None	S	S	I	S	S
			2010	None	None	S	S	S	S	S
			2012	None	None	S	S	I	S	S
			2015	*dfrA14*	IncFII, Col(pVC)	S	S	I	S	S
Asia 2[a] (n = 8)	ST2063 (n = 8)	Sri Lanka (n = 4)	2011	*bla*$_{TEM}$*-1B*	IncL/M(pOXA-48)	R	S	I	S	S
			2012	*aph(6)-Id, blaTEM-1B, QnrS1, sul3, tet(A), dfrA14*	IncX1	R	S	R	R	R
			2012	None	None	S	S	I	S	S
			2016	*QnrB19*	Col(pVC)	S	ND	I	S	ND
		Thailand (n = 2)	2015	None	Col(pVC)	S	ND	I	S	ND
			2016	None	Col(pVC)	S	ND	S	S	ND
		Not known (n = 1)	2014	None	Col(pVC)	S	S	I	S	S
		None (n = 1)	2015	None	Col(pVC)	S	ND	I	S	ND

BL: β-lactam antibiotics; CHL: chloramphenicol; I: intermediate resistance; ND: not done; QNL: quinolone; R: resistant; S: susceptible; ST: sequence type; SXT: trimethoprim-sulfamethoxazole; TET: tetracycline; WGS: whole genome sequencing.

[a] WGS clusters are named for the geographical regions where the majority of cases had travelled. Asia 1 and Asia 2 refer to two separate phylogenetic clusters we identified among the cases with travel history to Asia.

independent SeqSphere+ allele numbering nomenclature. The allelic profiles of the isolates were visualised as a neighbour-joining tree using the parameter 'pairwise ignoring missing values'.

Single nucleotide polymorphism-based reference mapping analysis

CSI Phylogeny, version 1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny/) was used to map sequences of our strains under investigation against *S*. Chester SRX992125 as a reference using the Burrows-Wheeler Aligner (BWA), call and filter SNPs through SAMtools using default parameters [13,14]. In MEGA6, the maximum likelihood method based on the Tamura-Nei model was used to infer phylogeny from the reference-based SNP calling, and a bootstrap consensus tree inferred from 1,000 replicates was also produced [15,16].

Discriminatory power

Calculations comparing the discriminatory power of MLST, cgMLST and reference mapping methods were performed using Simpson's index of diversity [17].

Antimicrobial susceptibility testing

All isolates were routinely tested for ampicillin, ciprofloxacin (CIP) and trimethoprim-sulfamethoxazole using the agar disk diffusion method. From 2016 onwards, quinolone resistance was inferred from pefloxacin resistance (n = 18). In addition, a selection of the isolates (n = 32) were screened for tetracycline (TET), chloramphenicol and nalidixic acid (NAL). Results were interpreted as sensitive (S), intermediate (I) or resistant (R) using the EUCAST clinical breakpoints, version 7.1 [18] when available, or based on epidemiological cut-off values of national zone distributions for CIP (S ≥ 33 mm and R < 30 mm), NAL (R < 16 mm) and TET (R < 17 mm) [19].

Characterisation of antimicrobial resistance determinants, single nucleotide polymorphisms associated with resistance and plasmids

The online tools ResFinder version 3.0, PlasmidFinder version 1.3 and pMLST version 1.4 available at the Center for Genomic Epidemiology (http://www.genomicepidemiology.org/) were used, respectively, for sequence-based identification of acquired resistance genes, known mutations conferring resistance and plasmid-borne genes, using assembled genomes obtained through SPAdes Genome Assembler version 3.0 (Algorithmic Biology Laboratory, St. Petersburg University, St. Petersburg, Russia) [20,21]. For PlasmidFinder, the threshold for minimum identity was set at 95% and for coverage at 80%. For ResFinder, the threshold for minimum identity was set at 90% and for coverage at 60%.

Results

Description of cases

Nine cases had no travel history outside Norway, the remaining 42 cases had a history of travel to southern Europe (n = 13), Africa (n = 16) or Asia (n = 12). The travel history of one case was unknown. The median age of the cases was 44 years (range: 7–86 years) and 26 of the 51 cases were male.

Seven of eight cases with a history of travel to Greece in the summer of 2016 were interviewed upon giving a sample positive for *S*. Chester; however, the only common exposure that was revealed was staying on Rhodes (7/7) with the majority staying in the city of Rhodes (6/7).

Serotypes

Conventional and sequence-predicted antigenic profiles were concordant, identifying the serotype Chester (4:e,h:e,n,x) for all isolates.

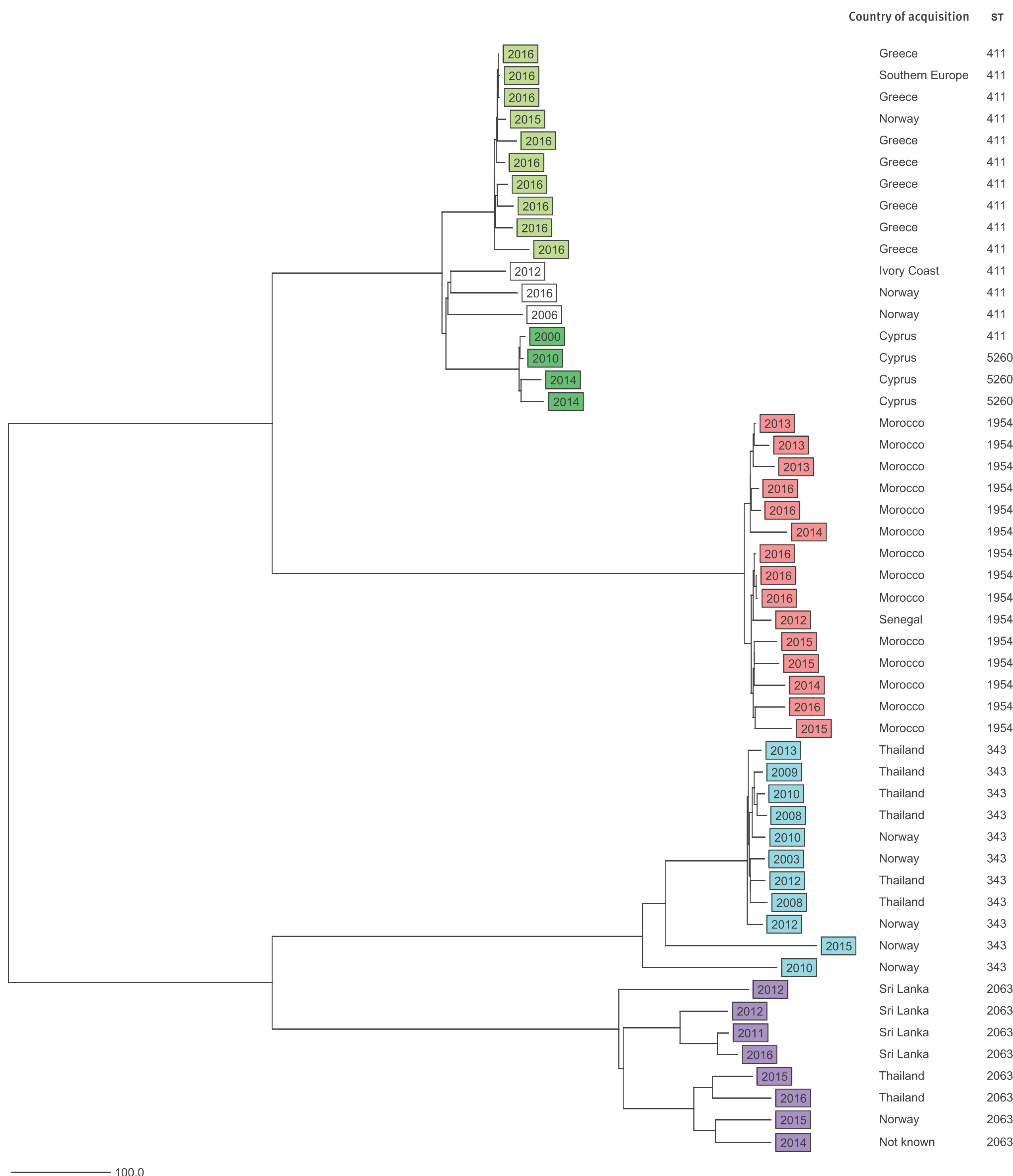
Multilocus sequence typing and core genome multilocus sequencing typing

The isolates represented five MLST sequence types (STs): ST1954 (n = 15), ST411 (n = 14), ST343 (n = 11), ST2063 (n = 8) as well as ST5260 (n = 3) which is a single-locus variant (SLV) of ST411 (Table).

All 51 *S*. Chester isolates had ≥ 98.6% good cgMLST targets (mean: 99.4%). Through cgMLST, based on 3,002 core genes and seven MLST genes, we identified four phylogenetic clusters separated by ≥ 719 allelic differences. These clusters were primarily associated with different geographical regions of acquisition: Europe, Africa and two separate clusters for Asia (Figure 2, Table). The European and Asian clusters included isolates from cases without reported history of travel outside Norway (Figure 2). Half of the cases in the Asia 1 cluster had a history of travel to Thailand (6/11) and in the Asia 2 cluster to Sri Lanka (4/8). The European cluster was further divided into Greek (n = 10) and Cypriot (n = 4) subclusters (Figure 2), while three isolates belonged to neither subcluster. The European subclusters were distanced from each other by ≥ 107 allelic differences. Within each of the two Asian clusters, the allelic differences between isolates were present in up to 8.2% (248/3,009) of the included genes, and within the European cluster in 4.2% (107/3,009) of the genes. There were fewer allelic differences between the isolates within the Greek (1.3%; 40/3,009) and Cypriot (0.9%; 27/3,009) subclusters and within the African cluster (1.5%; 45/3,009). Some identical isolates were also present: three isolates in the Greek subcluster were identical by cgMLST, as were two isolates in the African cluster. The years of isolation of *S*. Chester from the different clusters overlapped in time (Figure 2).

Figure 2

Neighbour-joining tree of *Salmonella* Chester isolates, based on 3,009 core genes included in core genome multilocus sequence typing, Norway, 2000–2016 (n = 51)



Nodes are labelled according to year of isolation. In addition, country of acquisition and seven-gene sequence type (ST) is noted. Scale bar shows absolute number of allelic differences. Green and white: European cluster (ST411 and ST5260); red: African cluster (ST1954); blue: Asian cluster 1 (ST343); lilac: Asian cluster 2 (ST2063). The analysis was performed in in Ridom SeqSphere+.

Single nucleotide polymorphism-based reference mapping analysis

Genome mapping phylogeny based on 14,176 SNPs revealed four main clusters that corresponded with the cgMLST results (Figure 3). The clusters were separated by ≥ 3,623 SNPs. Within the European cluster as a whole, there were 416 SNP differences. While within the Greek and Cypriot subclusters, there were ≤ 8 and ≤ 16 SNP differences, respectively, these subclusters were separated from the other three isolates within the European cluster by ≥ 170 SNPs. Within the African cluster, there were ≤ 51 SNP differences, while there were ≤ 601 and ≤ 852 SNP differences, respectively, within the Asia 1 and Asia 2 clusters.

Discriminatory power

The discriminatory power by Simpson's index of diversity was 0.78 for conventional MLST and 0.99 for cgMLST and genome mapping.

Antimicrobial resistance by phenotypic and genotypic characterisation

Overall, 16 of the 51 isolates were quinolone resistant by phenotypic testing. All 16 carried one or more quinolone resistance determinants: 10 carried the *qnrS1* gene and five carried the *qnrB19* gene, and the S83F SNP in *gyrA* was identified in two isolates, one of which also carried *qnrS1* (Table). All isolates resistant to chloramphenicol (6/6), trimethoprim-sulfamethoxazole (14/14) and tetracycline (9/9) carried known resistant determinants to these antimicrobials. The four isolates in this study that were resistant to β-lactams carried $bla_{TEM\text{-}1B}$ and were present in the two Asian clusters. Antimicrobial resistance varied between the clusters identified through WGS (Table). No isolates in the European cluster were fully resistant to any of the tested antimicrobials.

All isolates in the African cluster were resistant to at least one class of antibiotics. Most common in this cluster was trimethoprim-sulfamethoxazole resistance (n = 13), these resistant isolates carried both the *sul2* and *dfrA14* genes. Thirteen isolates in the African cluster were resistant to two or more classes of antibiotics. Twelve isolates in this cluster harboured the full set of resistance genes (*aph(3")-Ib (strA)*, *aph(6)-Id (strB)*, *sul2*, *tet(A)*, and/or *floR*) carried on the Tn3-like transposon that was identified in the outbreak cluster associated with travel to Morocco in a previous study [2].

Resistance to multiple antimicrobial agents was rare outside the African cluster. However, in the Asia 1 cluster, two isolates displayed resistance to both β-lactams and quinolones and carried the $bla_{TEM\text{-}1B}$ and *qnrS1* genes. In the Asia 2 cluster, one isolate carried the IncX1 plasmid and the *aph(6)-Id (strB)*, $bla_{TEM\text{-}1B}$, *qnrS1*, *sul3*, *tet(A)*, *dfrA14*genes, and was resistant to β-lactams, tetracycline, quinolone and trimethoprim-sulfamethoxazole. Across all the clusters, colicin bacteriocin-encoding Col-plasmids were carried by 25 of the 51 isolates. Plasmids of the incompatibility (Inc) types detected among the isolates were IncI1, IncI2, IncX1, IncX3(pEC14), IncFII, IncL/M(pOXA-48) and IncN. These were carried by 16 of the 51 isolates (Table). In the African cluster, six of 15 isolates carried the IncN-pST7 plasmid, which was not found in any of the other clusters.

Discussion

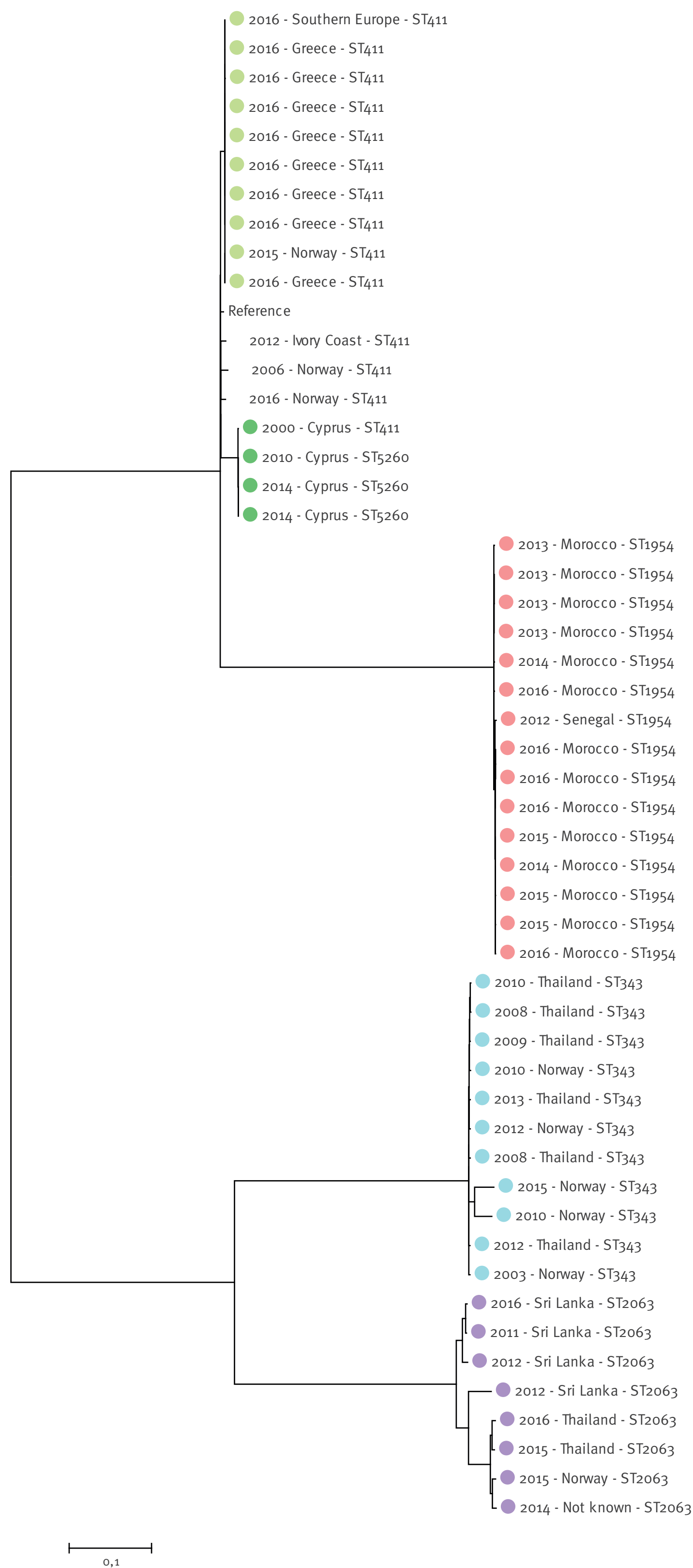
It is widely recognised that WGS-based methods offer higher resolution compared with conventional typing methods in distinguishing outbreak-associated isolates from sporadic ones [22,23]. For several enteropathogenic bacterial species, high concordance of results has been shown between cgMLST and reference mapping approaches, including *Salmonella* Enteritidis [24], *Listeria* [25] and *Enterococcus faecium* [26]. We observed similar concordance in our study, where the same four clusters were identified by both approaches. Both cgMLST- and SNP-based analyses identified the same Greek and Cypriot subclusters within the European cluster. The allelic and SNP differences observed in the two workflows also both confirmed the same phylogeny, where the European and African clusters were more similar to each other and more distant from the two Asian clusters. There was more internal diversity within the Asian clusters, compared with the internal diversity within the European subclusters and within the African cluster. The four main clusters identified through WGS displayed distinct STs based on conventional MLST for seven housekeeping genes, although the Cypriot subcluster within the European cluster included both ST411 and a novel SLV ST5260.

The discriminatory power of cgMLST- and SNP-based analyses was high and exceeded that of conventional MLST. The discriminatory power of cgMLST, combined with the ease of performing the analysis, the lower requirements of computational power and bioinformatics knowledge compared with a reference mapping SNP-based workflow, makes this an appropriate method for public health microbiology. Isolates can be analysed and compared with previously analysed isolates as they are received, which allows for continuous monitoring of potential outbreak clusters through gene-by-gene comparisons of a standardised cgMLST. In addition, because the scheme is standardised, the cgMLST EnteroBase can be used to describe the analysed isolates in a wider context by comparing them to other analysed isolates. For further analysis of clusters detected by cgMLST, SNP analysis can be performed for even greater resolution.

In this study, we used cgMLST to investigate genetic relatedness of *S.* Chester for surveillance and early outbreak detection and to compare the isolates from Norwegian patients who had travelled to Greece with isolates from patients with a history of travel to other geographic regions. As the majority of the *Salmonella* cases identified in Norway report a history of travel [27], our data offer some level of insight into the internationally circulating *S.*Chester strains.

Figure 3

Molecular phylogenetic analysis of *Salmonella* Chester isolates, based on single nucleotide polymorphism differences, Norway, 2000–2016 (n = 51)



The branch structure was confirmed by a bootstrap consensus tree inferred from 1,000 replicates. A total of 14,176 positions were present in the final dataset. S. Chester SRX992125 was used as a reference. Nodes are labelled according to year of isolation, country of acquisition and seven-gene sequence type (ST). Green and white: European cluster (ST411 and ST5260); red: African cluster (ST1954); blue: Asian cluster 1 (ST343); lilac: Asian cluster 2 (ST2063). The scale bar shows substitutions per site.

The travel history of our cases allowed us to identify geographical clusters, and our results also show that unrelated clusters, describing probable outbreaks, were overlapping in time. For example, the isolates from the African cluster, identified in the period between 2012 and 2016, were unrelated to the isolates associated with travel to Rhodes, Greece, which were identified in the summer of 2016. The isolates from cases who had travelled to Greece formed a separate group within the larger European cluster. The European cluster also included a smaller Cypriot subcluster, with four isolates from 2000, 2010, and 2014.

To further investigate the European cluster, we compared the ST411 isolates included in this study with the ST411 isolates deposited in EnteroBase. Of the 14 ST411 isolates included in our study, 10 clustered together with fewer than two allelic differences. They represent the Greek subcluster within the European ST411 cluster that we identified in our study, and the reported travel history of the 10 cases was Greece (n = 8), southern Europe (n = 1) and none (n = 1). In addition to our isolates, five ST411 isolates from the UK clustered within two allelic differences from our Greek subcluster, however, the travel history was unknown for the isolates from the UK.

While this and previous studies indicate that both SNP- and cgMLST-based WGS analysis can provide epidemiologically relevant microbiological information in the context of an outbreak investigation, it does not replace epidemiological information. In any outbreak investigation, microbiological and epidemiological data ideally complement each other in disentangling the outbreak, but microbiological data like these, especially when performed regularly as molecular surveillance, may alert to potential outbreaks that require epidemiological investigation. Although interviews were unable to confirm a source or common exposure for the cases travel-related to Rhodes, Greece, the WGS results of both the SNP-based and the cgMLST analysis give reason to believe that these cases constituted an outbreak. Surprisingly, our EPIS enquiry did not return any reports from similar findings elsewhere in Europe, although Rhodes is a holiday destination for many Europeans.

The isolates within the African cluster and within the Cypriot subcluster were genetically similar, although spanning several years, and isolates such as these should be flagged in WGS-based molecular surveillance for possible further investigation. In contrast, the two Asian clusters revealed through WGS in our study were geographically less contained, with cases reporting travel to one of several Asian countries or no travel abroad, and spanned a period of several years. They were also genetically more diverse and probably do not represent outbreak clusters, but rather a sample of the strain population that circulates in Asia and perhaps elsewhere.

As WGS is increasingly employed in public health microbiology to provide epidemiologically relevant information for outbreak investigations and surveillance, harmonised or standardised cut-offs for cluster definitions are needed and have already been proposed for some species [28]. Because of the inherent characteristics of the WGS analysis methods, we can expect that the SNP variation will be greater than the allelic differences in the same cluster, therefore the cut-offs must be adjusted not only to the species under investigation and possibly to subtype, serovar or serotype, but also to the WGS analysis approach. In addition, SNP-based results may differ from each other depending on trimming and pruning quality parameters defined in the SNP identification process, and some suggest that it may be impossible to define single cut-off values for outbreaks [14,29]. For cgMLST, a cut-off value for clusters would most probably need to take into account the number of core genes included in the analysis, and perhaps the cut-off could be a percentage of allelic differences rather than an absolute number. However, even for cgMLST, different assembly software could introduce some variation into the results, even when using the same sequencing chemistry.

The clusters associated with travel to Europe and Asia included cases without travel history. However, domestically acquired salmonellosis cases are rare in Norway [27]. These results therefore invite speculation on the possibility of secondary transmission to people in Norway from persons with travel history or through consumption of imported food items. Previous studies on salmonellosis have concluded that most cases are contracted through contaminated food, while person-to-person transmission is rare [30]. Inadvertent omission of travel details in connection with specimen collection or isolate submission is also a possibility that could explain these results.

In our study, two clusters were related to travel to Asia, one with just over half of the cases reporting travel to Thailand, and the other with travel history to Sri Lanka or Thailand. Geographical clusters were also identified in a previous study focusing on the multinational outbreak of *S.* Chester in Europe related to travel to Morocco in 2014 and 2015 [2], however, our study indicates that the outbreak may have been still ongoing in late 2016, as four cases belonging to the cluster and reporting travel to Morocco were identified in Norway in November 2016.

Antimicrobial resistance varied between the clusters. Aside from six intermediately quinolone-resistant isolates, all isolates in the European cluster were susceptible to all tested antibiotics, while some resistance was seen in the two Asian clusters. The three isolates displaying resistance to both quinolones and β-lactams and one isolate resistant to β-lactams were part of the two Asian clusters. The results for the isolates from the African cluster, where resistance to

antibiotics was frequent, are in agreement with prior knowledge about the ST1954 cluster [2]. It has been concluded that the use of antibiotics in treating non-severe *Salmonella* diarrhoea offers no clinical benefits and that antibiotics appear to increase adverse effects and may prolong the presence of *Salmonella* [31]. However, as antibiotic use plays an important role in the development of antibiotic resistance, this variation between clusters may reflect variation in the use of antimicrobials for humans and livestock between the originating regions. Estimating global consumption of antimicrobials in animals is challenging, but experts estimate that it will increase by 67% from 2010 to 2030 [32]. Studies describing findings of *S.* Chester from animal feed and faeces are available for two African countries. In a study of *Salmonella* in animal feed commercially produced in Namibia, *S.* Chester was the most commonly encountered serovar; however, resistance was rare. In a separate study, *S.* Chester isolates with intermediate resistance to streptomycin were discovered in poultry and cattle faeces in Burkina Faso [33,34].

A previous study by Fonteneau et al., focusing on the multinational *S.* Chester outbreak related to Morocco, found that isolates carrying the IncN-*qnrS1* plasmid appeared in 2014 [2]. In our material, isolates harbouring this plasmid were isolated already in 2013. In our study, these isolates were also ST1954 and originated from cases with history of travel to Morocco, which indicates that one of the sources in the multisource outbreak may have been active already then. As IncN plasmids are more commonly identified in isolates from animals than from humans, it has previously been suggested that the plasmids could have acquired the *qnrS1* gene in animals [35]. Six of our isolates harboured the same IncN-pST7 plasmid that was first reported in isolates connected to the Moroccan outbreak [2]. Two thirds of the resistant isolates in our study carried plasmids that have been linked with plasmid-mediated quinolone resistance [35]. All fully quinolone resistant isolates carried one of the *qnr* genes and/or point mutations known to confer resistance. The Inc plasmid types identified in our study were not confined to one WGS cluster, and the isolates in a cluster did not all carry the same plasmids.

A limitation of our study is the convenience sample of *S.* Chester isolates included in the analysis. However, we have attempted to mitigate this by including all isolates submitted to the National Reference Laboratory in the years from 2014 to 2016, and the additional isolates were selected to represent multiple years and a variety of geographical origins. A second limitation of our interpretation of the results is that we do at this point not have universal defined cut-offs of the number of SNP or allelic differences to determine clusters for *S.* Chester. However, we believe that in the future, as WGS continues to be used and more genomes become available in the public databases, our possibilities to determine exact cut-offs for defining a cluster will improve through sharing data such as those we obtained in this study.

Conclusion

WGS of *S.* Chester cases in Norway shows geographically distinct clusters associated with travel history of the patients and with varying antimicrobial susceptibility profiles between clusters. Although standardised cut-off values for relatedness as defined through WGS need more epidemiological validation and further data, our results indicate an outbreak of *S.* Chester in Norway during summer 2016. They further indicate that the outbreak was related to travel to Rhodes, Greece, and different from the simultaneous multicountry outbreak associated with travel to Morocco [2]. We recommend implementing cgMLST-based molecular surveillance for accurate and timely detection of future outbreaks for *S.* Chester and other *S. enterica* isolates.

Acknowledgements

We would like to thank Ina Haagesen and other technical staff at the Department of Bacteriology and at the Department of Molecular Biology at the Norwegian Institute of Public Health (NIPH) for their work. We thank the medical microbiological laboratories in Norway for collecting and submitting isolates to the NIPH. We also gratefully acknowledge local and international EUPHEM coordinators for guidance during this study and Silvia Herrera Leon for reviewing the manuscript.

Conflict of interest

LS is a co-investigator in an unrelated study, for which the National Institute for Health and Welfare, Finland, received research funding from GlaxoSmithKline Biologicals SA.

Authors’ contributions

UN, LTB and LS designed the study, performed sequence analysis and wrote the manuscript. NOH and KA were responsible for phenotypic and molecular laboratory methods. HL was responsible for the epidemiological investigations. All authors commented and agreed upon the final manuscript.

References

1. European Food Safety Authority (EFSA), European Centre for Disease Prevention and Control. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. EFSA Journal. 2017;15(12):5077.
2. Fonteneau L, Jourdan Da Silva N, Fabre L, Ashton P, Torpdahl M, Müller L, et al. Multinational outbreak of travel-related Salmonella Chester infections in Europe, summers 2014 and 2015. Euro Surveill. 2017;22(7):30463. https://doi.org/10.2807/1560-7917.ES.2017.22.7.30463 PMID: 28230522
3. European Food Safety Authority (EFSA), European Centre for Disease Prevention and Control. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2014. EFSA J. 2015;13(12):4329.
4. Taylor J, Galanis E, Wilcott L, Hoang L, Stone J, Ekkert J, et al. An outbreak of salmonella chester infection in Canada: rare serotype, uncommon exposure, and unusual population demographic facilitate rapid identification of food vehicle. J Food Prot. 2012;75(4):738-42. https://doi.org/10.4315/0362-028X.JFP-11-408 PMID: 22488063
5. Centers for Disease Control and Prevention (CDC). Multistate outbreak of salmonella chester infections associated with

frozen meals -- 18 states, 2010. MMWR Morb Mortal Wkly Rep. 2013;62(48):979-82. PMID: 24304829
6. Guo Z, Su Ch, Huang J, Niu J. A food-borne outbreak of gastroenteritis caused by different Salmonella serotypes in 2 universities in Xiamen, Fujian, China, in 2012. Jpn J Infect Dis. 2015;68(3):187-91. https://doi.org/10.7883/yoken.JJID.2014.235 PMID: 25672350
7. O'Grady KA, Krause V. An outbreak of salmonellosis linked to a marine turtle. Southeast Asian J Trop Med Public Health. 1999;30(2):324-7. PMID: 10774704
8. OzFoodNet Working Group. Burden and causes of foodborne disease in Australia: Annual report of the OzFoodNet network, 2005. Commun Dis Intell Q Rep. 2006;30(3):278-300. PMID: 17120483
9. Eurobarometer F. 392. Preferences of Europeans toward tourism. Brussels: European Commission; 2014. Available from: http://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_392_en.pdf
10. Grimont PAD, Weill FX. Antigenic formulae of the Salmonella serovars. Paris: Institut Pasteur; 2007. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.231.3561&rep=rep1&type=pdf
11. Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, et al. Salmonella serotype determination utilizing high-throughput genome sequencing data. J Clin Microbiol. 2015;53(5):1685-92. https://doi.org/10.1128/JCM.00323-15 PMID: 25762776
12. Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of Salmonella. PLoS Genet. 2018;14(4):e1007261. https://doi.org/10.1371/journal.pgen.1007261 PMID: 29621240
13. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One. 2014;9(8):e104984. https://doi.org/10.1371/journal.pone.0104984 PMID: 25110940
14. Ahrenfeldt J, Skaarup C, Hasman H, Pedersen AG, Aarestrup FM, Lund O. Bacterial whole genome-based phylogeny: construction of a new benchmarking dataset and assessment of some existing methods. BMC Genomics. 2017;18(1):19. https://doi.org/10.1186/s12864-016-3407-6 PMID: 28056767
15. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512-26. PMID: 8336541
16. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. https://doi.org/10.1093/molbev/mst197 PMID: 24132122
17. Hunter PR. Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol. 1990;28(9):1903-5. PMID: 2229371
18. The European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 7.1. Växjö: EUCAST; 2017. Available from: http://www.eucast.org.
19. NORM/NORM-VET. 2016. Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway. Tromso/Oslo: University Hospital of North Norway and Norwegian Veterinary Institute; 2017. Available from: https://unn.no/Documents/Kompetansetjenester,%20-sentre%20og%20fagr%C3%A5d/NORM%20-%20Norsk%20overv%C3%A5kingssystem%20for%20antibiotikaresistens%20hos%20mikrober/Rapporter/NORM%20NORM-VET%202016.pdf
20. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903. https://doi.org/10.1128/AAC.02412-14 PMID: 24777092
21. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640-4. https://doi.org/10.1093/jac/dks261 PMID: 22782487
22. Bekal S, Berry C, Reimer AR, Van Domselaar G, Beaudry G, Fournier E, et al. Usefulness of high-quality core genome single-nucleotide variant analysis for subtyping the highly clonal and the most prevalent Salmonella enterica serovar Heidelberg clone in the context of outbreak investigations. J Clin Microbiol. 2016;54(2):289-95. https://doi.org/10.1128/JCM.02200-15 PMID: 26582830
23. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of Salmonella enterica. PLoS One. 2014;9(2):e87991. https://doi.org/10.1371/journal.pone.0087991 PMID: 24505344
24. Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ. Comparative analysis of core genome MLST and SNP typing within a European Salmonella serovar Enteritidis outbreak. Int J Food Microbiol. 2018;274:1-11. https://doi.org/10.1016/j.ijfoodmicro.2018.02.023 PMID: 29574242
25. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, et al. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of Listeria monocytogenes. J Clin Microbiol. 2015;53(9):2869-76. https://doi.org/10.1128/JCM.01193-15 PMID: 26135865
26. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, et al. Core genome multilocus sequence typing scheme for high- resolution typing of Enterococcus faecium. J Clin Microbiol. 2015;53(12):3788-97. https://doi.org/10.1128/JCM.01946-15 PMID: 26400782
27. MacDonald E, White R, Mexia R, Bruun T, Kapperud G, Brandal LT, et al. The role of domestic reservoirs in domestically acquired Salmonella infections in Norway: epidemiology of salmonellosis, 2000-2015, and results of a national prospective case-control study, 2010-2012. Epidemiol Infect. 2018;15:1-8. https://doi.org/10.1017/S0950268818002911 PMID: 30428947
28. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect. 2018;24(4):350-4. https://doi.org/10.1016/j.cmi.2017.12.016 PMID: 29309930
29. Saltykova A, Wuyts V, Mattheus W, Bertrand S, Roosens NHC, Marchal K, et al. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to Salmonella enterica serotype Typhimurium and serotype 1,4,[5],12:i. PLoS One. 2018;13(2):e0192504. https://doi.org/10.1371/journal.pone.0192504 PMID: 29408896
30. Horwitz MA, Pollard RA, Merson MH, Martin SM. A large outbreak of foodborne salmonellosis on the Navajo Nation Indian Reservation, epidemiology and secondary transmission. Am J Public Health. 1977;67(11):1071-6. https://doi.org/10.2105/AJPH.67.11.1071 PMID: 911019
31. Sirinavin S, Garner P. Antibiotics for treating salmonella gut infections. Cochrane Database Syst Rev. 2000; (2):CD001167. PMID: 10796610
32. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, et al. Global trends in antimicrobial use in food animals. Proc Natl Acad Sci USA. 2015;112(18):5649-54. https://doi.org/10.1073/pnas.1503141112 PMID: 25792457
33. Kagambèga A, Lienemann T, Aulu L, Traoré AS, Barro N, Siitonen A, et al. Prevalence and characterization of Salmonella enterica from the feces of cattle, poultry, swine and hedgehogs in Burkina Faso and their comparison to human Salmonella isolates. BMC Microbiol. 2013;13(1):253. https://doi.org/10.1186/1471-2180-13-253 PMID: 24215206
34. Shilangale RP, Di Giannatale E, Chimwamurombe PM, Kaaya GP. Prevalence and antimicrobial resistance pattern of Salmonella in animal feed produced in Namibia. Vet Ital. 2012;48(2):125-32. PMID: 22718330
35. Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009;53(6):2227-38. https://doi.org/10.1128/AAC.01707-08 PMID: 19307361

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Research

Air-conditioner cooling towers as complex reservoirs and continuous source of *Legionella pneumophila* infection evidenced by a genomic analysis study in 2017, Switzerland

Daniel Wüthrich[1,2,3], Sylvia Gautsch[4], Ruth Spieler-Denz[5], Olivier Dubuis[6], Valeria Gaia[7], Jacob Moran-Gilad[8,9], Vladimira Hinic[1], Helena MB Seth-Smith[1,2,3], Christian H. Nickel[10], Sarah Tschudin-Sutter[11], Stefano Bassetti[12], Monika Haenggi[13], Peter Brodmann[4], Simon Fuchs[5], Adrian Egli[1,2]

1. Division of Clinical Microbiology, University Hospital Basel, Basel, Switzerland
2. Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland
3. Swiss Institute of Bioinformatics, Basel, Switzerland
4. State Laboratory Basel-City, Basel, Switzerland
5. Department of Health, Medical Services, Canton of Basel-Stadt, Basel, Switzerland
6. Viollier, Allschwil, Switzerland
7. National Reference Center for Legionella, Department of Laboratory medicine, Ente Ospedaliero Cantonale, Bellinzona, Switzerland
8. Department of Health Systems Management, School of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
9. Public Health Services, Ministry of Health, Jerusalem, Israel
10. Division of Emergency Medicine, University Hospital Basel, Basel, Switzerland
11. Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland
12. Division of Internal Medicine, University Hospital Basel, Basel, Switzerland
13. Department of Health, Medical Services, Canton of Basel-Country, Liestal, Switzerland

Correspondence: Daniel Wüthrich (daniel.wuethrich@usb.ch)

Citation style for this article:
Wüthrich Daniel, Gautsch Sylvia, Spieler-Denz Ruth, Dubuis Olivier, Gaia Valeria, Moran-Gilad Jacob, Hinic Vladimira, Seth-Smith Helena MB, Nickel Christian H., Tschudin-Sutter Sarah, Bassetti Stefano, Haenggi Monika, Brodmann Peter, Fuchs Simon, Egli Adrian. Air-conditioner cooling towers as complex reservoirs and continuous source of Legionella pneumophila infection evidenced by a genomic analysis study in 2017, Switzerland. Euro Surveill. 2019;24(4):pii=1800192. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800192

Article submitted on 15 Apr 2018 / accepted on 07 Nov 2018 / published on 24 Jan 2019

Introduction: Water supply and air-conditioner cooling towers (ACCT) are potential sources of *Legionella pneumophila* infection in people. During outbreaks, traditional typing methods cannot sufficiently segregate *L. pneumophila* strains to reliably trace back transmissions to these artificial water systems. Moreover, because multiple *L. pneumophila* strains may be present within these systems, methods to adequately distinguish strains are needed. Whole genome sequencing (WGS) and core genome multilocus sequence typing (cgMLST), with their higher resolution are helpful in this respect. In summer 2017, the health administration of the city of Basel detected an increase of *L. pneumophila*infections compared with previous months, signalling an outbreak. **Aim:** We aimed to identify *L. pneumophila* strains populating suspected environmental sources of the outbreak, and to assess the relations between these strains and clinical outbreak strains. **Methods:** An epidemiological and WGS-based microbiological investigation was performed, involving isolates from the local water supply and two ACCTs (n = 60), clinical outbreak and non-outbreak related isolates from 2017 (n = 8) and historic isolates from 2003–2016 (n = 26). **Results:** In both ACCTs, multiple strains were found. Phylogenetic analysis of the ACCT isolates showed a diversity of a few hundred allelic differences in cgMLST. Furthermore, two isolates from one ACCT showed no allelic differences to three clinical isolates from 2017. Five clinical isolates collected in the Basel area in the last decade were also identical in cgMLST to recent isolates from the two ACCTs. **Conclusion:** Current outbreak-related and historic isolates were linked to ACCTs, which form a complex environmental habitat where strains are conserved over years.

Introduction

Legionella pneumophila (Lp) causes Legionnaires' disease (LD), a severe infection of the respiratory tract. LD was first described in 1976 after an outbreak at an American legion convention due to a contaminated air-conditioning system [1]. In that outbreak, 182 persons were infected and 29 (16%) died [1]. Since then, Lp has been considered an important threat to public health. The European Legionnaires' disease Surveillance network (ELDSNet) reported, that between 2011 and 2015 across 29 European countries a total of 30,532 LD cases were documented, whereas the incidence rose from 0.97 (2011) to 1.30 (2015) per 100,000 inhabitants.

Most LD cases are community-acquired and affect people aged 50 years or older, with mortality rates around 10% [2]. In Switzerland during 2017, 492 cases of LD were reported [3] with an incidence of 5.81 per 100,000 inhabitants.

Infections with Lp are acquired via inhalation of contaminated aerosolised water [4]. Various environmental sources are known, such as showers [5,6], hot tubs, fountains, dishwashers [7], hot water tanks, larger plumbing systems [8] and air-conditioner cooling towers (ACCT) [9-12].

A given environmental source can host several types of Lp strains, which, in some cases, can enter amoeba biofilms [4] leading to low mutation rates and a high conservation of genomic diversity. As a consequence, traditional typing methods such as serotyping, pulsed-field electrophoresis (PFGE), and sequence-based typing (SBT) do not provide sufficient resolution to trace outbreaks to individual sources. In addition, certain Lp clonal complexes of clinical relevance (e.g. ST1) are spread worldwide, and respective isolates are so similar, that SBT cannot distinguish them [13]. This renders SBT insufficient for typing Lp for public health purposes. On the other hand, different isolates of a specific clonal complex may have a limited number of single nt polymorphisms (SNPs) (e.g. ST1: 121 SNPs), which are detectable by whole genome sequencing (WGS) to allow their discrimination.

The ability of WGS to deliver complete genomic information [14], thereby conferring higher-resolution, has made it the gold standard for typing Lp isolates. Moreover, investigations of LD incidents in a fast and automatic manner have recently been facilitated by a core genome multilocus sequence typing (cgMLST) scheme based on WGS data [15]. Beside single outbreak investigations, WGS-based typing data also support comparison across studies [16]. Nevertheless, many recent reports on Lp using WGS have mainly focused on single outbreaks [5,8,12,17,18]. These studies also did not assess the complexity of environmental sources in great detail, whereby the sampling strategy of the environmental isolates and the diversity of strains in the sources remain unclear.

Based on epidemiological evidence, ACCTs are suspected to be a considerable source of outbreaks [17,19-21], yet the *Legionella* populations within have not been thoroughly described. To clarify the transmission mechanism of Lp, which in turn guides appropriate control measures, it is important to understand the environmental complexity of Lp populations (e.g. genomic diversity, exchange between populations) and relate this to data from outbreak-related clinical isolates. The goal of this study was to extend our knowledge of the role of environmental Lp sources, such as ACCTs and water supply, during an outbreak, or over a prolonged time period. Therefore, we studied clinical isolates from the city of Basel and surrounding areas during an outbreak in 2017 and compared these to isolates originating from water pipes and ACCTs by applying WGS. We also sequenced clinical isolates that were collected since 2003. With these data we attempt to identify links between Lp populations within ACCTs, and outbreak-related and historical clinical isolates.

Methods

Setting

In Switzerland, all positive Lp cases have to be reported to the federal office of health by law [3], which is followed by an environmental risk assessment. Briefly, cases clinically suspected of respiratory tract infection get screened using a urinary *Legionella* antigen testing according to the manufacturer (BinaxNOW from Alere, which detects serotypes 1–14 or Sofia Legionella FIA from Quidel, which detects serotype 1; San Diego, United States). In the case of a positive Lp result, the treating physician is contacted to report the result and send respiratory material for culture-based detection and subsequent typing of the Lp isolate.

We cultured 34 strains from humans. Four cultured isolates (isolate ID: NMB001740, NMB001739, NMB001863, NMB001758) of Lp serotype 1 obtained in the time period of the outbreak and the specific city district associated with the outbreak were available for WGS analysis. As non-outbreak controls, we included four serotype 1 isolates from the same time period, but different geographical areas including the neighbouring cantons (n = 3) and another city district of Basel (n = 1). Furthermore, we included 26 historic isolates collected between 2003 and 2016 in the canton of Basel-city and the neighbouring cantons. Additionally, we used 60 Lp isolates from the local water supply chain and ACCTs within the area of the outbreak (up to 29 isolates per location). The details of the samples are listed in Supplementary Table S1.

Origins, culture and serogroup identification of human isolates

Respiratory materials such as sputum, tracheal secretion and bronchioalveolar lavages, were cultured for a maximum of 10 days at 36 °C under 5% CO_2 using buffered media with polymyxin B, anisomycin and alpha-ketoglutarate (BMPA from Thermoscientific, Reinach, Switzerland) and standard 5% sheep blood agar (bioMérieux, Lyon, France). Culture plates were daily checked for growth and suspected colonies were identified using matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (Microflex system, Bruker, Bremen, Germany). Lp isolates were further separated into serogroup 1 or 2–14 (Legionella latex test, Oxoid (Pratteln, Switzerland)). The historic isolates were obtained from the strain collection of the University Hospital Basel and respective serogroups were determined in the same way.

Origins, culture and serogroup identification of environmental isolates

Water samples (1,000 mL) from suspected environmental sources (tap water sources/plumbing systems, ACCTs) were filtered and cultured directly and after filtration without treatment, after acid treatment and after heat treatment following the International Standard ISO 11731:2017 'Water quality – Enumeration of *Legionella*'. The isolates were cultured aerobically for a maximum of 10 days at 37 °C using the selective media buffered charcoal yeast extract agar with polymyxin B, anisomycin and cephamandol (BMPA from Oxoid, Pratteln, Switzerland), MWY (buffered charcoal yeast extract agar with glycin, polymyxin B, anisomycin, vancomycin, bromothymol blue and bromocresol purple; Oxoid, Pratteln, Switzerland) and GVPC (buffered charcoal yeast extract agar with glycin, vancomycin, polymyxin B and cycloheximide; Oxoid, Pratteln, Switzerland). Culture plates were checked every 2–3 days for growth and suspected colonies were identified by subculture on buffered charcoal yeast extract agar (BCYE-agar; Oxoid, Pratteln, Switzerland) with L-cysteine and on standard 5% sheep blood agar (bioMérieux, Lyon, France). Isolates showing no growth on cysteine-free blood agar were considered as *Legionella* and further identified by agglutination and separated into serogroup 1 or 2–14 (Legionella latex test, Oxoid (Pratteln, Switzerland)). Finally, colony forming units of Lp per mL and per 1,000mL of water sample were determined.

Whole genome sequencing of bacterial isolates and bioinformatic analysis

From both clinical and environmental isolates, we included each morphotype to WGS analysis. DNA from cultured isolates was extracted using a robotic system (EZ1 Advanced XL, Qiagen (Venlo, Netherlands)). WGS sequencing was performed using a MiSeq Illumina platform (accredited with ISO 17025 norm) with 2x 300nt paired-end sequencing as previously described [22]. The resulting reads were de novo assembled using Unicycler [23] (version 0.4.4) and the assemblies (assembly statistics are listed in Supplementary Table S2) used for cgMLST analysis performed with Ridom SeqSphere Software (version 4.1.9) using the recently published cgMLST scheme [15]. All isolates had at least a mean coverage of 90-fold. All genomes sequenced for this study were submitted to GenBank (see accession numbers Supplementary Table S2).

All available Lp genome assemblies were downloaded from the National Center for Biotechnology Information (NCBI, December 2017, 539 genomes). The assemblies were re-annotated using Prokka (version 1.12) [24] for consistency, and phylogenetic analysis that was based on the core genome alignment was performed using Roary (version 3.11.2) [25] and FastTree (version 2.1) [26]. The phylogenetic tree was visualised using iTOL [27]. Whole genome comparison (SNP-calling) was performed using BWA (version 0.7.17) [28] and Pilon (version 1.22) [29].

Results

Description of *Legionella pneumophila* outbreak in Basel 2017

In 2017, the weekly number of LD cases in the city of Basel appeared to increase from May to August (Supplementary Figure S1). In this city, the overall incidence per 100,000 inhabitants increased from 4.66 to 15.02 between 2016 and 2017 [3] (Supplementary Figure S2). Although, no active case finding strategy was developed, the health administration of the city of Basel performed a detailed epidemiological investigation using a standardised questionnaire to assess potential risk factors for *Legionella* exposure for all infected patients (Supplement S1, Supplement S2). Based on the investigation results, including the place of residence of the patients, a spatial and temporal cluster of Lp serotype 1 infected patients in a particular city district was found. A secondary investigation with more specific questions about epidemiological risk factors and places visited was performed. Thereby, the area and particular exposures could be even further specified. Interestingly, ACCTs were found in the vicinity of some of the patients' homes.

Whole genome sequencing typing of human isolates

The WGS-based cgMLST comparison showed that three of the four putative outbreak isolates had the same cgMLST type (cluster type 228), with no allelic differences (0/1521). Therefore, these patients were infected with the same strain. The other five clinical (including the one outbreak isolate, and the four non-outbreak isolates) isolates from 2017 (cluster types are listed in Supplementary Table S1) showed more than 90 allelic differences to the cluster of three samples, indicating that these patients were infected with other strains.

Investigation of environmental sources

In order to identify a possible source of infection for the three patient isolates sharing the same cgMLST type, we sampled water sources from plumbing systems in close proximity to their respective place of residence. We identified four different locations contaminated with Lp (Supplementary Table S1). Colonies with various morphotypes were selected. The investigated locations contained Lp serotypes 1 and 2–14. Because patients had been tested with a serotype 1 specific urinary antibody test in 2017 only serotype 1 clinical isolates were available for comparison.

Based on the epidemiological assessment of the outbreak cluster patients, we suspected eight ACCTs as possible sources of *Legionella*. As there is no cooling tower registry for the city of Basel, we used the epidemiological risk assessment to identify the most likely towers in close distance for the sampling. Material from two of these eight sampled ACCTs yielded growth of Lp. In the two ACCTs various morphotypes could be detected on the culture plates, including serotypes 1 and

2–14, all of which were included into the WGS-analysis. Quantitative analysis of *Legionella*in the water samples from these contaminated ACCTs reached up to 5.8 million colony forming units per litre (Supplementary Table S1), which reflects a high pathogen density. The water flow and aerosolisation associated with functions of an ACCT are shown in Supplementary Figure S3.

Whole genome sequencing analysis of clinical and environmental isolates

WGS was performed on 37 isolates from the two contaminated ACCTs and 23 isolates from the four plumbing systems contaminated with Lp (Supplementary Table S1). Isolates were selected based on differing morphotypes from the different isolation sites. In addition, we included 26 historic clinical isolates from the strain collection of the University Hospital Basel, collected since 2003 from the city of Basel and surrounding area.

A total of 94 clinical and environmental isolates were analysed using cgMLST. The overall diversity throughout all isolates was very high, covering more than one thousand allelic differences (Figure), producing 13 closely related complexes (≤10 allelic differences) and also 15 strains without close relation to other isolates (Figure).

Our first analysis focused on the environmental samples, which were found within ten complexes (Figure). ACCT-derived isolates can be found within six different complexes, while all environmental isolates recovered from tap water and plumbing sources (Figure) were found in four different clusters. Complexes 1 and 2 contain isolates from the two ACCTs sites. Most interestingly, isolates within complex 1 originated from both ACCTs, including some isolates from both ACCTs with no allelic differences.

The comparison of the environmental and clinical samples showed that the three identical clinical outbreak isolates are closely associated with two isolates from a single ACCT (Figure, 'complex 9'), showing no allelic differences in the cgMLST analysis. This analysis was complemented by a whole genome based variant calling approach for increased typing resolution. This approach revealed a variability of only 5 SNPs within that cluster, further highlighting the close relatedness (data not shown). We also analysed all serogroup 1 strains of the ACCT using SBT that showed that all strains were ST36 (Philadelphia). Furthermore, we also found that five historic isolates, sampled between 2003 and 2011, were within complex 1, and showed no allelic differences to four environmental isolates from the ACCTs (Figure, 'complex 1'). Therefore, we concluded that this environmental strain, recovered during the current investigation, has been causing infections over the past decades. In total, we observed that 12 clinical isolates (historic and 2017) had 10 or fewer allelic differences compared with the closest related environmental isolate.

Nevertheless, not all historic or current clinical isolates could be linked to the sampled environmental isolates. Interestingly, we found that of these 22 clinical isolates (17 historic and five from 2017) nine are found in three complexes (Figure, complexes 4,6,7). Especially interesting is complex 4, as it contains five clinical isolates from 2009 to 2017. The remaining 13 clinical isolates are not closely related to any other isolate. To investigate potential origins of these 22 patient strains without connection to environmental isolates, we accessed 539 Lp genomes from the public NCBI database, reflecting a global strain collection. We performed phylogenetic analysis of all sequenced strains from Basel and the genomes from NCBI that were isolated in 17 different countries (clinical and environmental samples). In order to handle the high number of genomes (n = 633), we compared the strains using a core genome alignment-based phylogeny. The analysis showed that all 22 strains without links to environmental isolates are closely related to isolates from other European countries (Supplementary Figure S4).

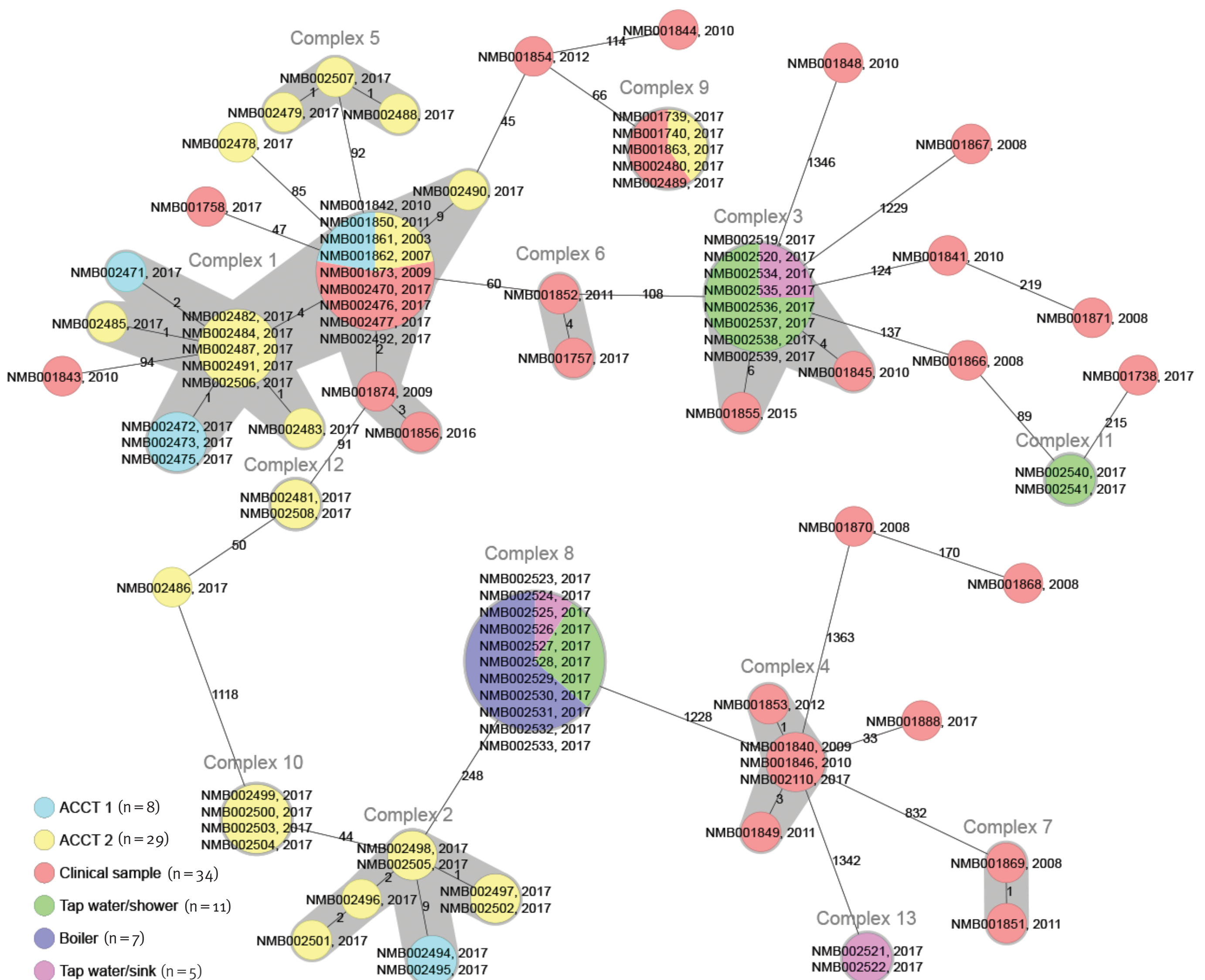
Discussion

In this study, we have shown, based on WGS and cgMLST analysis, that clinical isolates associated with the outbreak of 2017 in the city of Basel are genetically related to ACCT-derived isolates. This finding supports that ACCTs can act as a source of *Legionella* infection, as suspected in previous studies [19,20]. An important finding of our study is the broad genetic diversity of environmental isolates across the city. Although the isolates sampled from two ACCTs were found to be very closely related (complex 1 and 2), the findings clearly highlight the need to sample a broad range of environmental reservoirs in an outbreak setting in order to identify the causal source. Due to the diversity within these environmental reservoirs, we believe that shotgun metagenomics [30] could provide more information than WGS on selected isolates, as the latter might overlook important strains. However, this approach would necessitate the use of appropriate and maybe newly developed bioinformatics tools that allow the differentiation of strains in metagenomics samples [31,32].

We have demonstrated that identical environmental isolates can be found in different sampling locations, potentially indicating a complex environmental network. As there was no direct water pipe connection between the two contaminated ACCTs in this study that are ca 500 m apart, our current assumption is that the release of contaminated aerosols not only leads to human exposure, but also facilitates the exchange of *Legionella* populations between ACCTs. Some previous studies have attempted to characterise Lp populations in ACCTs. In 103 water samples from 50 ACCTs collected over five years in Turkey (1996–2000), relatively stable serotype distributions with 44% serotype 1 have been described [33]. Another study used 16S sequencing to study the *Legionella* species dynamics within cooling towers and found that Lp can

Figure

Allelic differences between the *Legionella pneumophila* strains recovered in clinical and environmental isolates, Switzerland, 2003–2017 (n = 94 isolates)



ACCT: air-conditioner cooling towers; cgMLST: core genome multilocus sequence typing.

The circles represent sequenced strains. The circles are coloured according to the isolation source. The numbers next to the lines connecting two circles indicate the number of allelic differences. Strains that have no allelic differences are listed in the same circle. Strains with 10 or less allelic differences are clustered into complexes and connected with a grey background. The analysis is based on the published cgMLST scheme [15] for *Legionella pneumophila* using 1,521 allelic loci. The last four numbers next to the isolate ID indicate the isolation year

outcompete other *Legionella* species [30]. However, to date no high-resolution analysis of Lp within ACCTs has been conducted. Our findings highlight the potential of (i) a complex environmental network and (ii) suggest that decontaminated ACCTs (the decontamination automatics were defect in the observed ACCTs) can be potentially recolonised by contaminated aerosols from other ACCTs. This information could be used to influence the design of ACCTs (Supplementary Figure S3) and strategies in the control of potential outbreak sources [34,35]. Our study already had a real-life effect, as for the two contaminated ACCTs, the maintenance procedure for decontamination was corrected after our findings.

The diversity of isolates within one environmental *Legionella* population, as shown by the WGS data, is also remarkable. We found isolates from the same populations that are separated by more than one thousand allelic differences. Interestingly, only environmental isolates from complex 1, 3 and 9 were connected to clinical samples (Figure). The cgMLST results indicates that subclones of the same ST (e.g. ST36) seem to have enhanced potential for causing infection, as out of the 15 cluster types that we found in the environmental

samples, only cluster type 177 and 228 also comprised the clinical samples.

The inclusion of previously collected isolates from the strain collection of the University Hospital Basel allowed us to increase the sample size, and also to link historic cases to environmental contamination. The transmission from ACCTs appears not to be a rare event that is limited to the outbreak from 2017. We were able to connect some clinical isolates found in ACCTs to clinical infections that occurred almost a decade apart and the strains can still be found in the ACCT (Figure, 'complex 1'). However, this is not limited to only one event, we found several cases of closely related clinical isolates that were isolated in different years (Figure). We concluded that these are conserved *Legionella* strains in environmental sources, that lead to infections over several years and that these environmental sources form a complex network. This is in agreement with another study, where the same strains were found over several years [8]. We assume strains are conserved over the years in biofilms [4]. Unfortunately, no historical environmental samples were available to test our hypothesis.

While our findings provide more insight into potential links between complex environmental Lp reservoirs and LD, this study has several limitations. First, we only had a limited number of isolates available, in particular isolates were not obtained from all outbreak-related patients. Although a total of 94 isolates were included, the study would certainly have further benefitted from a higher sampling density of environmental and clinical isolates. Often clinical isolates cannot be collected, as patients with a positive antigen test in urine samples will receive treatment and no culture isolation from respiratory material is performed. Clearly, physicians should be aware of the importance of *Legionella*culture and WGS-based typing for public health reasons. In addition, the sensitivity of culture-based methods for *Legionella* detection is somewhat limited [36]. Another limitation was that unfortunately, we did not have historical samples from the environment that could match historical clinical samples. Finally, we were only able to sample two ACCT sites, although the exchange of strains between both systems could be documented, more systems should be sampled and analysed in the future.

In conclusion, we showed that contaminated ACCTs are an important threat to public health. WGS played a crucial role in this study, as it allowed the high-resolution typing and therefore demonstrated the value of this technique in clinical microbiology. In particular, the potential that environmental systems can form a complex network without having a direct water supply connection is an important finding. Finally, we have shown that strains are conserved and cause infections over decades.

Acknowledgements

We want to thank Magdalena Schneider, Christine Kiessling, Elisabeth Schultheiss, Rosa-Maria Vesco and Clarisse Straub for the DNA extraction, library preparations and sequencing of the bacterial isolates.

This study was partially funded by the City health administration of Basel, Switzerland. AE received a salary grant from the Swiss National Science Foundation (PZ00P3_154709 / 1).

Conflict of interest

None declared.

Authors' contributions

DW, SG, SF and AE designed and coordinated the study; SG, OD, VH, PB and AE provided clinical and environmental isolates; DW analysed the WGS data; VG provided SBT data; DW, SG, VH, RSD, OD, VG, JMG, HSS, CN, STS, SB, MH, PB, SF and AE wrote the manuscript.

References

1. Winn WC Jr. Legionnaires disease: historical perspective. Clin Microbiol Rev. 1988;1(1):60-81. https://doi.org/10.1128/CMR.1.1.60 PMID: 3060246
2. Beauté JThe European Legionnaires' Disease Surveillance Network. Legionnaires' disease in Europe, 2011 to 2015. Euro Surveill. 2017;22(27). https://doi.org/10.2807/1560-7917.ES.2017.22.27.30566 PMID: 28703097
3. Federal Office of Public Health (FOPH). Liebefeld: FOPH. Available from: https://www.bag.admin.ch
4. Cunha BA, Burillo A, Bouza E. Legionnaires' disease. Lancet. 2016;387(10016):376-85. https://doi.org/10.1016/S0140-6736(15)60078-2 PMID: 26231463
5. Schjørring S, Stegger M, Kjelsø C, Lilje B, Bangsborg JM, Petersen RF, et al. ESCMID Study Group for Legionella Infections (ESGLI). Genomic investigation of a suspected outbreak of Legionella pneumophila ST82 reveals undetected heterogeneity by the present gold-standard methods, Denmark, July to November 2014. Euro Surveill. 2017;22(25):30558. https://doi.org/10.2807/1560-7917.ES.2017.22.25.30558 PMID: 28662761
6. Laganà P, Gambuzza ME, Delia S. Legionella risk assessment in cruise ships and ferries. Ann Agric Environ Med. 2017;24(2):276-82. https://doi.org/10.26444/aaem/74717 PMID: 28664708
7. Yoshida M, Furuya N, Hosokawa N, Kanamori H, Kaku M, Koide M, et al. Legionella pneumophila contamination of hospital dishwashers. Am J Infect Control. 2018;46(8):943-5. https://doi.org/10.1016/j.ajic.2018.01.024 PMID: 29502885
8. Rosendahl Madsen AM, Holm A, Jensen TG, Knudsen E, Lundgaard H, Skov MN, et al. Whole-genome sequencing for identification of the source in hospital-acquired Legionnaires' disease. J Hosp Infect. 2017;96(4):392-5. https://doi.org/10.1016/j.jhin.2017.04.020 PMID: 28622979
9. Lucas KD, Wheeler C, McLendon P, Leistikow BN, Mohle-Boetani JC. Outbreak of Legionnaires' disease associated with cooling towers at a California state prison, 2015. Epidemiol Infect. 2018;146(3):297-302. https://doi.org/10.1017/S0950268818000110 PMID: 29386076
10. Llewellyn AC, Lucas CE, Roberts SE, Brown EW, Nayak BS, Raphael BH, et al. Distribution of Legionella and bacterial community composition among regionally diverse US cooling towers. PLoS One. 2017;12(12):e0189937. https://doi.org/10.1371/journal.pone.0189937 PMID: 29261791
11. Timms VJ, Rockett R, Bachmann NL, Martinez E, Wang Q, Chen SC, et al. Genome sequencing links persistent outbreak of legionellosis in Sydney (New South Wales, Australia) to an emerging clone of Legionella pneumophila sequence type 211. Appl Environ Microbiol. 2018;84(5):e02020-17. https://doi.org/10.1128/AEM.02020-17 PMID: 29247056
12. Lapierre P, Nazarian E, Zhu Y, Wroblewski D, Saylors A, Passaretti T, et al. Legionnaires' disease outbreak caused by endemic strain of Legionella pneumophila, New York,

New York, USA, 2015. Emerg Infect Dis. 2017;23(11):1784-91. https://doi.org/10.3201/eid2311.170308 PMID: 29047425

13. David S, Rusniok C, Mentasti M, Gomez-Valero L, Harris SR, Lechat P, et al. Multiple major disease-associated clones of Legionella pneumophila have emerged recently and independently. Genome Res. 2016;26(11):1555-64. https://doi.org/10.1101/gr.209536.116 PMID: 27662900
14. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J, Laurent F, et al. . Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18(4):20380. https://doi.org/10.2807/ese.18.04.20380-en PMID: 23369389
15. Moran-Gilad J, Prior K, Yakunin E, Harrison TG, Underwood A, Lazarovitch T, et al. Design and application of a core genome multilocus sequence typing scheme for investigation of Legionnaires' disease incidents. Euro Surveill. 2015;20(28):1-9. https://doi.org/10.2807/1560-7917.ES2015.20.28.21186 PMID: 26212142
16. 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. https://doi.org/10.1128/JCM.06094-11 PMID: 22238442
17. Shivaji T, Sousa Pinto C, San-Bento A, Oliveira Serra LA, Valente J, Machado J, et al. A large community outbreak of Legionnaires disease in Vila Franca de Xira, Portugal, October to November 2014. Euro Surveill. 2014;19(50):20991. https://doi.org/10.2807/1560-7917.ES2014.19.50.20991 PMID: 25597540
18. David S, Afshar B, Mentasti M, Ginevra C, Podglajen I, Harris SR, et al. Seeding and establishment of legionella pneumophila in hospitals: Implications for genomic investigations of nosocomial legionnaires' disease. Clin Infect Dis. 2017;64(9):1251-9. https://doi.org/10.1093/cid/cix153 PMID: 28203790
19. Garbe PL, Davis BJ, Weisfeld JS, Markowitz L, Miner P, Garrity F, et al. Nosocomial Legionnaires' disease. Epidemiologic demonstration of cooling towers as a source. JAMA. 1985;254(4):521-4. https://doi.org/10.1001/jama.1985.03360040075028 PMID: 4009880
20. Bosch X. Legionnaire's outbreak in Spanish town may be largest ever. Lancet. 2001;358(9277):220. https://doi.org/10.1016/S0140-6736(01)05465-4 PMID: 11476856
21. McAdam PR, Vander Broek CW, Lindsay DS, Ward MJ, Hanson MF, Gillies M, et al. Gene flow in environmental Legionella pneumophila leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. Genome Biol. 2014;15(11):504. PMID: 25370747
22. Piso RJ, Käch R, Pop R, Zillig D, Schibli U, Bassetti S, et al. A cross-sectional study of colonization rates with Methicillin-Resistant Staphylococcus aureus (MRSA) and Extended-Spectrum Beta-Lactamase (ESBL) and carbapenemase-producing Enterobacteriaceae in four Swiss refugee centres. PLoS One. 2017;12(1):e0170251. https://doi.org/10.1371/journal.pone.0170251 PMID: 28085966
23. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol. 2017;13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595 PMID: 28594827
24. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9. https://doi.org/10.1093/bioinformatics/btu153 PMID: 24642063
25. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. https://doi.org/10.1093/bioinformatics/btv421 PMID: 26198102
26. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823
27. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8. https://doi.org/10.1093/bioinformatics/btl529 PMID: 17050570
28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168
29. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. https://doi.org/10.1371/journal.pone.0112963 PMID: 25409509
30. Wéry N, Bru-Adan V, Minervini C, Delgénes JP, Garrelly L, Godon JJ. Dynamics of Legionella spp. and bacterial populations during the proliferation of L. pneumophila in a cooling tower facility. Appl Environ Microbiol. 2008;74(10):3030-7. https://doi.org/10.1128/AEM.02760-07 PMID: 18390683
31. Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat Methods. 2016;13(5):435-8. https://doi.org/10.1038/nmeth.3802 PMID: 26999001
32. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 2017;27(4):626-38. https://doi.org/10.1101/gr.216242.116 PMID: 28167665
33. Türetgen I, Sungur EI, Cotuk A. Enumeration of Legionella pneumophila in cooling tower water systems. Environ Monit Assess. 2005;100(1-3):53-8. https://doi.org/10.1007/s10661-005-7058-3 PMID: 15727299
34. Mouchtouri VA, Goutziana G, Kremastinou J, Hadjichristodoulou C. Legionella species colonization in cooling towers: risk factors and assessment of control measures. Am J Infect Control. 2010;38(1):50-5. https://doi.org/10.1016/j.ajic.2009.04.285 PMID: 19699013
35. Yu VL. Cooling towers and legionellosis: a conundrum with proposed solutions. Int J Hyg Environ Health. 2008;211(3-4):229-34. https://doi.org/10.1016/j.ijheh.2008.02.003 PMID: 18406666
36. Reller BL, Weinstein MP, Murdoch DR. Diagnosis of Legionella infection. Clin Infect Dis. 2003;36(1):64-9. https://doi.org/10.1086/345529 PMID: 12491204

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Review

Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018

Belén Rodríguez-Sánchez[1,2,3], Emilia Cercenado[1,2,4], Alix T. Coste[5], Gilbert Greub[3,5,6]
1. Department of Clinical Microbiology and Infectious Diseases, Hospital General Universitario Gregorio Marañón, Madrid, Spain
2. Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
3. These authors contributed equally to this work
4. Department of Medicine, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain
5. Institute of Microbiology, University Hospital of Lausanne, Lausanne, Switzerland
6. Infectious Diseases Service, University Hospital of Lausanne, Lausanne, Switzerland

Correspondence: Belén Rodriguez-Sanchez (mbelen.rodriguez@iisgm.com)

Citation style for this article:
Rodríguez-Sánchez Belén, Cercenado Emilia, Coste Alix T., Greub Gilbert. Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018. Euro Surveill. 2019;24(4):pii=1800193. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800193

Article submitted on 15 Apr 2018 / accepted on 04 Jan 2019 / published on 24 Jan 2019

Introduction: MALDI-TOF MS represents a new technological era for microbiology laboratories. Improved sample processing and expanded databases have facilitated rapid and direct identification of microorganisms from some clinical samples. Automated analysis of protein spectra from different microbial populations is emerging as a potential tool for epidemiological studies and is expected to impact public health. **Aim:** To demonstrate how implementation of MALDI-TOF MS has changed the way microorganisms are identified, how its applications keep increasing and its impact on public health and hospital hygiene. **Methods:** A review of the available literature in PubMED, published between 2009 and 2018, was carried out. **Results:** Of 9,709 articles retrieved, 108 were included in the review. They show that rapid identification of a growing number of microorganisms using MALDI-TOF MS has allowed for optimisation of patient management through prompt initiation of directed antimicrobial treatment. The diagnosis of Gram-negative bacteraemia directly from blood culture pellets has positively impacted antibiotic streamlining, length of hospital stay and costs per patient. The flexibility of MALDI-TOF MS has encouraged new forms of use, such as detecting antibiotic resistance mechanisms (e.g. carbapenemases), which provides valuable information in a reduced turnaround time. MALDI-TOF MS has also been successfully applied to bacterial typing. **Conclusions:** MALDI-TOF MS is a powerful method for protein analysis. The increase in speed of pathogen detection enables improvement of antimicrobial therapy, infection prevention and control measures leading to positive impact on public health. For antibiotic susceptibility testing and bacterial typing, it represents a rapid alternative to time-consuming conventional techniques.

Introduction

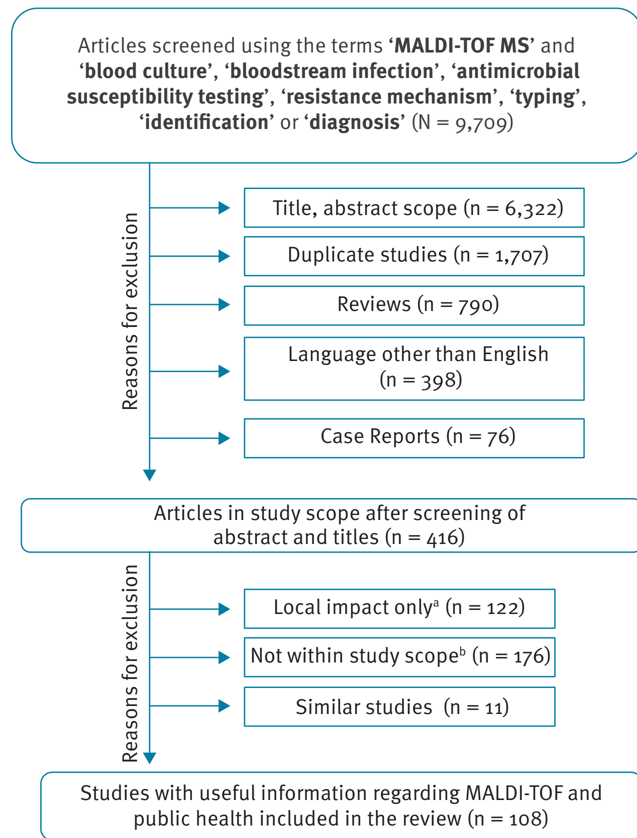
During the past 10 years, Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) has changed microbiology routine practice by allowing timely and cost-effective identification of different microorganisms, not only from pure culture but also directly from clinical samples [1-3]. Indeed, faster microbial identification allows for earlier antibiotic streamlining, due to the accurate identification provided for important groups of microorganisms that can be managed with directed antibiotic treatment, as demonstrated when MALDI-TOF MS was applied to bacterial identification directly from blood culture pellets [4,5]. MALDI-TOF MS has also been applied to determine antimicrobial susceptibility patterns, and has produced reliable same-day results; this is a major advantage, as routine antimicrobial susceptibility testing (AST) analyses typically need overnight incubation [6].

MALDI-TOF MS has also emerged as a diagnostic tool for bacterial typing, which could help to detect nosocomial outbreaks, with a putative beneficial impact on disease control and patient safety [7,8]. Hospital hygiene may also benefit from early identification of some emerging and clinically relevant pathogens [9]; in this context, the rapid identification of pathogens, even at the subspecies or serotype level, may positively impact the time until patient isolation and the prompt initiation of the appropriate drug therapy. In some circumstances, such as the recent *Mycobacterium chimaera* outbreak, early identification of atypical mycobacteria would also prove useful to detect such case clusters [10].

Altogether, in this review we aim to demonstrate that MALDI-TOF MS represents a versatile diagnostic technology with great potential to improve the identification of microorganisms and to impact public health

Figure 1

Flowchart of the literature retrieved and retained in review of MALDI-TOF MS use in public health and hospital hygiene, 2018



[a] Regional or very local microbiological problems indicated by a limited number of samples (n <10).

[b] MALDI-TOF MS was used as an identification tool, but its performance was not the objective of the study.

by providing important information for optimised antimicrobial stewardship and disease prevention and control.

Methods

A review of the available scientific literature was carried out. We searched the United States (US) National Institutes of Health's National Library of Medicine PubMed database for articles published in English between January 2009 and October 2018, using the terms 'MALDI-TOF', 'blood culture', 'bloodstream infection', 'antibiotic susceptibility testing', 'resistance mechanism', 'typing', 'highly pathogenic microorganisms', 'identification' and 'diagnosis'.

Reference lists from published articles were also screened to find more literature on the topic. In addition, reports from the European Centre for Disease Prevention and Control (ECDC) were consulted to identify outbreaks and public health issues where MALDI-TOF MS was applied to detect the causing pathogen (https://ecdc.europa.eu/en/threats-and-outbreaks).

The articles identified in the search were screened based on the information included in their titles and abstracts. Studies with a scope other than the application of MALDI-TOF MS on public health and hospital hygiene issues, as well as duplicate studies, were excluded. Case reports, studies acknowledging regional or very local microbiological problems (indicated by a very limited number of samples (n <10) and reviews were also excluded, though their reference lists were checked for related literature. Subsequently, the remaining articles were each assigned to an author for review, according to their area of expertise (including direct application of MALDI-TOF MS on blood cultures (GG), detection of resistance mechanisms (AC, EC), identification of public health-relevant microorganisms and typing with MALDI-TOF MS (BRS, EC).

Results

Literature selection

A total of 9,709 articles were found using the selected keywords. Based on the information in the titles and abstracts, 6,322 studies were out of scope and were therefore excluded. Among the remaining 3,387 papers, 1,707 appeared in the search results more than once and 790 reviews did not contribute new content because they reproduced results previously obtained by other authors in a different geographical area; these were also excluded. Articles written in languages other than English (n = 398) and 76 case reports referring to a very limited number of samples or patients were excluded as well (Figure 1).

At this stage, the remaining number of references was 416. During a second review, studies acknowledging regional or very local microbiological problems (n = 122) and those where MALDI-TOF MS was used as an identification tool but its performance was not the objective of the study (n = 176) were also excluded. In addition, in order to fall within the maximum number of references for publication, only the most recent articles showing similar design and results were included; all relevant articles are covered here (Figure 1).

Finally, 108 original articles demonstrating proof of concept, as well as a clear impact on microbiology and the microbiology laboratory praxis regarding the application of MALDI-TOF MS, were included in this review.

Implementation and clinical impact of performing MALDI-TOF MS on blood culture pellets

One of the most impactful uses of MALDI-TOF MS is its ability to identify microorganisms grown in blood cultures [11]. This application has shown to provide reliable identification of possible contaminants and disease-causing pathogens, as well as to reduce

turnaround time (TAT) to final identification, since overnight culture on agar media is not necessary [4,12].

Already in 2010, several authors proposed to prepare a bacterial pellet from positive blood cultures in order to fasten pathogen identification [13-15]. Since then, a variety of protocols have been used that reported identification of the aetiological agent of bacteraemia in 70–80% of cases, with accuracy greater than 99% (reviewed in [16]).

In these protocols, sample preparation aims at concentrating the microorganisms present in the blood culture by using differential centrifugation and washing steps. Then, the pellet can be spotted directly on the MALDI target for identification [17] or be submitted to a protein extraction procedure [18]. The use of the SepsiTyper kit (Bruker Daltonics, Billerica, Massachusetts, US) has also been reported for this purpose [19]; its performance was shown to be similar to the direct and protein-extraction methods, but it provided superior results for yeasts identification. These results were supported by several studies [3,20]. In-house methods using different reagents also reported improved identification of yeasts and fungi in blood cultures [21,22]. Croxatto et al. developed an ammonium chloride-based approach to lyse red blood cells and obtain a clean bacterial pellet [23]. A short incubation step right after the blood culture bottle is flagged positive was also tested successfully [24]. The detection of beta-lactamases and carbapenemases using MALDI-TOF MS can also be applied on the obtained pellet. The procedure is detailed further down in this review [6,25].

So far, the drawback of MALDI-TOF MS directly on blood culture detected so far is the inability to identify all bacteria in a polymicrobial infection [17]. It has been overcome by the development of several AST approaches coupled to the identification of the causing pathogen (Figure 2).

The advantage of MALDI-TOF MS over conventional methods is that it offers a reliable identification of the pathogen and AST results can be obtained within one working shift in a rapid and inexpensive manner [26]. The clinical impact of the implementation of MALDI-TOF MS on blood cultures has been measured; in a study by Clerc et al. [5], MALDI-TOF MS allowed the adjustment of antibiotic treatment in 35.1% of the bacteraemia cases analysed. Without considering the centrifugation steps, the cost was calculated to be ca EUR 1.43 per sample tested, whereas the hospital stay was shown to be reduced by ca 2 days, depending on the patient type and the appropriateness of patient management [26,27]. Due to common use of carbapenems for septic shock at their study site, Clerc et al. observed antibiotic streamlining more often than broadening, with routine MALDI-TOF MS applied to blood culture pellets having a clear positive impact on reducing the usage of carbapenems and other broad-spectrum antibiotics [5].

A recent prospective study confirmed that identification of the aetiological agent of bacteraemia by MALDI-TOF MS led to a shorter time to adequate antibiotic treatment [28]. In this study, patients with *ampC*-positive, Gram-negative bacteriaemia rapidly identified by MALDI-TOF MS were optimally treated within 48 hours.

Thus, in several centres the implementation of MALDI-TOF MS for the routine identification of microorganisms directly from blood culture pellets has shown that it may significantly impact the streamlining of antibiotics, with a likely positive impact on the antibiotic resistance rate.

Identification and typing of epidemiologically relevant pathogens

The high specificity shown by MALDI-TOF MS in different studies encouraged researchers to further analyse the protein spectra obtained for identification of different microorganisms and to attempt comparison between subpopulations.

Food-borne pathogens

Discrimination at the subspecies or even serotype level has been researched for different bacterial genera of public health interest. For *Salmonella* spp., the finding of specific peaks that allow genus-, species-, subspecies- and even serotype-level discrimination has been described by Dieckmann et al. [29]. Using a decision tree based on the presence/absence of specific peaks, corresponding mainly to ribosomal proteins, the authors achieved correct identification of the most commonly encountered *S. enterica* subsp. *enterica* serotypes with 100% sensitivity and specificity. More recently, a study using similar peaks as serotype biomarkers and ad hoc software allowed 94% of correct *S. enterica* subsp. *enterica* serotype assignment using a set of 12 species-specific peaks [30]. The authors reported up to 96% correct serotype identification when the software reduced the number of biomarkers used to 10, with no impact on the specificity of the analysis. It is noteworthy that both studies used a whole-cell approach for serotyping, which requires a limited number of reagents and short TAT. The manual process of peak analysis can be more time-consuming and requires trained personnel. However, this requirement can be avoided by implementing specific software for peak analysis. The use of free software such as MALDIQuant [31] allows the simultaneous analysis of many spectra, with the necessity of a trained bioinformatician as the only drawback.

Further important food-borne pathogens that have been successfully subtyped with MALDI-TOF MS are Shiga-toxin producing *Enterobacteriaceae* [9]. The analysis of peak profiles yielded two important biomarkers that allowed correct identification of 103 of 104 *Escherichia coli* O104:H4 isolates from an outbreak that took place in northern Germany [9]. The implementation of MALDI-TOF MS from isolates spotted directly on a MALDI target plate or after a formic acid/acetonitrile extraction

renders this methodology very rapid, since the protein spectra can be obtained within minutes.

This approach has also allowed the discrimination of *Listeria monocytogenes*, a pathogen associated with a high mortality rate (20–30%) [32]. Beyond correct species-level identification of *L. monocytogenes* after culture conditions standardisation, the analysis of the protein spectra has allowed the source tracking of *L. monocytogenes* isolates from dairy sources [33] and the correct serotype assignment from clinical samples [34]. In addition, *L. monocytogenes* subtypes can be discriminated using the automated MALDI Biotyper (MBT) subtyping module developed by Bruker Daltonics [35].

Clostridium difficile

The implementation of MALDI-TOF MS for typing of *Clostridium difficile* has yielded successful results [36]. High molecular weight proteins from 500 isolates were analysed and high correlation with PCR ribotypes (89.0%) was reported. The availability of this easy-to-perform typing method allows rapid and accurate screening of outbreak-related *C. difficile* clones and helps epidemiologists and public health professionals to follow and control putative outbreaks.

Respiratory pathogens

Several respiratory pathogens of public health importance have been shown to be reliably identified using MALDI-TOF MS. *Legionella* spp. was identified from environmental samples in two different hospitals in a rapid and reliable manner [37,38].

Attempts to discriminate *Streptococcus pneumoniae* from the members of the *S. mitis* complex have yielded a panel of specific marker peaks that allow species assignment to *S. pneumoniae* isolates and the most common non-pneumococcal species (*S. mitis* and *S. oralis*) [39,40]. Compared with the culture from suspected isolates in the presence of an optochin disk, this MALDI-TOF MS application allows a reduction in TAT and laboratory costs [39].

Another group of important respiratory pathogens are the members of the *Mycobacterium* genus. MALDI-TOF MS cannot differentiate among the species comprising the *Mycobacterium tuberculosis* complex. Nevertheless, its implementation for the identification of non-tuberculous mycobacteria (NTM) has been useful for evaluating the clinical significance of the microorganism recovered by culture of various clinical samples. Around 60 NTM species have been shown to act as opportunistic human pathogens causing pulmonary disease with symptoms similar to tuberculosis lymphadenitis in children associated with *M. avium* and *M. scrofulaceum*, as well as skin diseases and disseminated infections in immunocompromised patients [41,42]. In this scenario, MALDI-TOF MS has shown to provide reliable species-level identification in almost 100% of the cases [43] and the sample processing methods available are easy to apply, require little hands-on time and are widely standardised [44]. However, MALDI-TOF MS applied to NTM grown on liquid medium exhibited a low sensitivity [45]. This drawback has been overcome, however, by some authors using an improved bead-based method for cell disruption. The implementation of this method reduced the TAT up to 2–3 weeks [46]. Closely related NTM species are often identified by MALDI-TOF MS at a complex level. However, Fangous et al. developed an algorithm that allowed the accurate discrimination between three subspecies within the *Mycobacterium abscessus* complex, namely *M. absces*sus subsp. *abscessus*, *M. abscessus* subsp. *massiliense* and *M. abscessus* subsp. *bolletii* [47]. The algorithm was based on the presence/absence of five specific peaks that correlated with the three subspecies. The discrimination of the subspecies within the *M. abscessus* complex was demonstrated as well by Kehrman et al. using principal component analysis [48]. In both cases, the discrimination of the subspecies was accurate and allowed for improved patient management due to the different antibiotic susceptibility pattern of each member of the *M. abscessus* complex. More recently, Pranada et al. have achieved a highly robust and accurate discrimination between *M. intracellulare* and *M. chimaera* by peak analysis [10]. Their approach supports the use of MALDI-TOF MS for the accurate discrimination of NTM isolates associated with heater/cooler devices used for extracorporeal cardiopulmonary support, an important issue in hospital hygiene and infection prevention [49].

Biosafety level Risk Group 3 pathogens

Highly pathogenic microorganisms are a major concern for their potential to be used as bioterrorism agents. The identification of Risk Group 3 bacterial pathogens with MALDI-TOF MS was assessed by different groups [50-52]. The authors reportedly showed no identification of these microorganisms when proprietary databases were employed. However, the use of the Security Relevant reference library, developed by Bruker Daltonics, allowed between 52.5–77.0% correct species assignment, although misidentifications with neighbour species were also reported [52]. The rate of correct species assignment reached the totality of the isolates tested only (i) when expanded with in-house libraries and/or (ii) when improved software for spectra analysis were used [51].

Recently, the US Centers for Disease Control and Prevention (CDC) collaborated with Bruker Daltonics in the construction of an expanded library for Risk Group 3 pathogens. This database can be accessed online (https://microbenet.cdc.gov/).

Finally, MALDI-TOF MS was able to identify the emerging pathogen *Candida auris*. The Biotyper updated Research Use Only (RUO) database already contains nine reference spectra from this pathogen, which allowed the discrimination from *C. haemuloni* without using an expanded library (data not shown).

Antimicrobial susceptibility detection using MALDI-TOF

Even without performing AST, the identification of microorganisms by MALDI-TOF MS impacts antimicrobial stewardship since the common susceptibility pattern of the identified microorganism can be largely deduced. This information can already be partially obtained by direct examination of the sample after performing a Gram staining, but MALDI-TOF MS goes one step further by giving at least the genera of the microorganism. Concerning Gram-negative rods, identification of group 3 *Enterobacterales* or a *Stenotrophomonas* spp. isolate, for example, will modify the antibacterial stewardship. It is the same for Gram-positive cocci and the possibility to distinguish *Enterococcus faecium* from *E. faecalis*, for example.

Considering the continuous emergence of acquired antibiotic/antifungal drug resistance, the need for same-day, full AST results become urgent. From this perspective, several studies have investigated the use of MALDI-TOF MS to perform AST. MALDI-TOF AST assays were first developed to detect specific peaks of resistant strains by peak picking approaches [53-55]. Most of these studies, however, concern detection of drug hydrolysis/modification (reviewed in [56]). Recently, some MALDI-TOF MS assays aimed at detecting drug resistance independently of the biological mechanism, evaluating the growth of a microorganism in the presence of a given drug [57-59].

The peak picking approaches

The first MALDI-TOF AST study was performed on *Staphylococcus aureus* to detect meticillin resistance [60]. Comparing the lists of peaks, some peaks specific for meticillin-resistant *S. aureus* (MRSA) and meticillin-susceptible *S. aureus* (MSSA) strains were identified. Further studies were then performed on larger sets of strains and on averaged spectra obtained from several replicates for a given strain. Cluster analysis was performed on the obtained peak list to discriminate MRSA from MSSA strains [61]. Interestingly, some authors of the first study also demonstrated that the cluster analysis result is modified depending on the growth media [62]. In contrast, Bernardo et al. showed that peak profiles were very stable regardless of the growth medium used. However, this study failed to define a clear peak signature for MRSA [63].

Other groups performing peak picking could discriminate between teicoplanin-susceptible vs -resistant staphylococci by analysing peak lists of laboratory-engineered mutant strains [64]. More recently, vancomycin intermediate-resistant *Staphylococcus aureus* (VISA) and vancomycin-susceptible *Staphylococcus aureus* (VSSA) could be discriminated by the identification of 22 relevant peaks using linear regression analysis, followed by a principal component analysis (PCA) on the identified peaks [65]. Once again, the influence of the growth medium on the obtained spectra was documented [65]. In 2018, Asakura et al. [66] further developed the machine learning approach initiated by Mather et al. to discriminate profiles of VISA among MRSA and heterogeneous VISA (hVISA) among MSSA, with 99% sensitivity for both. They also developed an 'all-in-one' online software publicly available to analyse in-house spectra [66]. The same approach was used earlier to discriminate *cfiA*-positive and *cfiA*-negative *Bacteroides fragilis* [67].

Since antimicrobial resistance is often due to the production of enzymes modifying the microorganism metabolism or degrading the drug, some MALDI-TOF MS studies developed assays to identify peaks corresponding to such enzymes. Studies were then performed to detect disappearance of peaks corresponding to *E. coli* or *Klebsiella pneumoniae* porins in spectra of strains with high resistance against beta-lactams [53]. This approach allows discrimination between carbapenemase expression and loss of porin expression conjugated with AmpC or extended Spectrum Beta-Lactamase (ESBL). Other groups were able to identify the peak of beta-lactamase at 29,000 m/z in ampicillin-resistant *E. coli* [55]. Concerning the detection of *B. fragilis* resistant to carbapenems, peaks specific to the IS insertion upstream of the *cfiA* gene were determined and a MBT subtyping module from Bruker Daltonics' was released to detect them [35].

Meticillin resistance in *Staphylococcus* is due to the acquisition of the *mecA or mecC* gene. The *mecA* gene is often acquired in parallel to the *psm-mec* gene coding for a toxin. Rhoads et al. specifically detected a peak near 2,415 m/z (± 2.00 m/z) that correlated with meticillin resistance (*mecA* carriage) in a series of consecutive staphylococcal blood culture isolates; this peak was present in 37% of the MRSA and 0% of MSSA strains [68]. Recently, Bruker Daltonic's MBT subtyping module included the detection of a peak corresponding to the PSM-mec peptide in *Staphylococcus aureus* spectra [35].

Detection of drug hydrolysis

The most important outcome of using MALDI-TOF AST so far was the detection of antimicrobial modifications, either quinolones acetylation or beta-lactam ring hydrolysis, leading to mass shift of 43 Da and 18 Da, respectively [56,69,70]. Beta-lactam ring hydrolysis is directly followed by a decarboxylation corresponding to a minus 44 Da shift. Thus, beta-lactamase hydrolysis rather appears as a minus 26 Da shift [71].

In 2011, Sparbier et al. established an interesting table of detected peaks for each type of beta-lactams before and after hydrolysis decarboxylation, in presence or absence of salts [70]. They then correlated the calculated data with measured data on strains incubated for 3 hours with the different drugs. By visual peak analysing, they obtained the same susceptibility and resistance results as routine AST methods. Further studies aimed to detect ESBL *Enterobacterales* through third-generation cephalosporins degradation [70,72]. To

quantify the hydrolysis, Jung et al. calculated the logarithm of the hydrolysed/non-hydrolysed peaks. This so-called LogRQ ratio discriminates drug susceptibility with 100% sensitivity and 91.5% specificity [72], even if criteria to interpret the ratio were not clearly defined. In a subsequent study, De Carolis et al. calculated the average intensity of the hydrolysed vs non-hydrolysed peaks, and compared them with the positive and negative control peaks [73]. Both studies investigated the possibility to detect enzyme activity directly in the blood culture pellet and obtained sensitivity and specificity of ca 87% and 98%, respectively.

The majority of the MALDI-TOF AST studies, however, focused on carbapenemases detection, as they represent a challenge for hospital hygiene as an emergent antimicrobial resistance mechanism. Several studies successfully detected carbapenemase-producing bacteria using different carbapenems as substrate, such as ertapenem [74,75], imipenem [76,77] and meropenem [78,79]. However, OXA48 carbapenemase in *Enterobacteriaceae* or imipenemases in *Pseudomonas aeruginosa* remain difficult to detect [80,81]. The addition of a bicarbonate buffer improved hydrolysis by *Enterobacteriaceae* of ertapenem and meropenem, but not imipenem [77,82]. Similarly, addition of zinc ion (Zn^{2+}) conserves activity of zinc-dependent *P. aeruginosa* imipenemases [83]. However, Rotova et al. highlighted a slightly better efficacy of meropenem supplemented with sodium dodecyl sulfate (SDS) and bicarbonate to detect *Enterobacteriaceae* and *Pseudomonas* carbapenemases than imipenem plus Zn^{2+} [84].

All these MALDI-TOF MS detections of drug modifications have lead, so far, to the development of the MBT STAR-BL software (Bruker Daltonics) and to one carbapenemase detection kit called MBT STAR-Carba Kit (Bruker Daltonics). Recent studies demonstrated the efficacy of this software with a concomitant identification and detection of ESBL or carbapenemase in around 1.5–5.2 hours, instead of 12–48 hours, with conventional routine protocols [85,86].

Detection of global spectra modifications in the presence of a drug

One promising use of MALDI-TOF AST consists of comparing spectra obtained from microorganisms in absence or presence of an antimicrobial agent. This approach was first developed in 2009 to discriminate between fluconazole-susceptible and -resistant *C. albicans* strains [87]. Authors compared spectra of *Candida* cells incubated in increasing concentrations of fluconazole. The minimal profile change concentration (MPCC) was determined as the lower concentration of drug needed to observe modification in the *C. albicans* spectra. Like for classical minimum inhibitory concentration (MIC), breakpoints were defined and then susceptible or resistant phenotypes could be easily determined after a few hours of incubation in fluconazole [87], allowing same-day antifungal susceptibility testing results. De Carolis et al. and Vella et al. further developed spectra comparison, performing a cross correlation index (CCI) matrix with spectra obtained in only three conditions: no drug, breakpoint and high concentration with a reduced 3-hour incubation [88,89]. They also adapted the method to echinocandins [90], other triazoles and other *Candida* species [91]. The overall agreement of the MALDI-TOF AST with the Clinical and Laboratory Standards Institute (CLSI) method ranged from 54–97%, depending on the species and the drug [92].

Comparison of spectra in the presence of a drug was also developed to determine bacterial resistance. It consists of a semiquantitative evaluation of the growth measuring intensity of different peaks in presence/absence of a drug following an internal standard [93]. First assays were performed using meropenem and *Klebsiella* strains [57]. Best results were obtained after 1 hour of incubation, reaching 97.3% sensitivity and 93.5% specificity. Like for the yeast assays described earlier, breakpoints were determined to distinguish susceptible from resistant strains. This approach was enlarged to cefotaxime, piperacillin-tazobactam, ciprofloxacin and gentamicin, other *Enterobacteriaceae* and non-fermenting Gram-negative rods, and it was adapted to blood culture samples [58]. The same methodology was tested for mycobacteria AST and allowed shortening of the TAT to one week for the NTM [59]. An example of such methodology is the MBT-ASTRA kit (MALDI BioTyper Antibiotic Susceptibility Test Rapid Assay, Bruker Daltonics), a promising tool for low-cost, same-day AST results on a wide range of pathogens and drugs [94-97].

Other applications of MALDI-TOF in public health

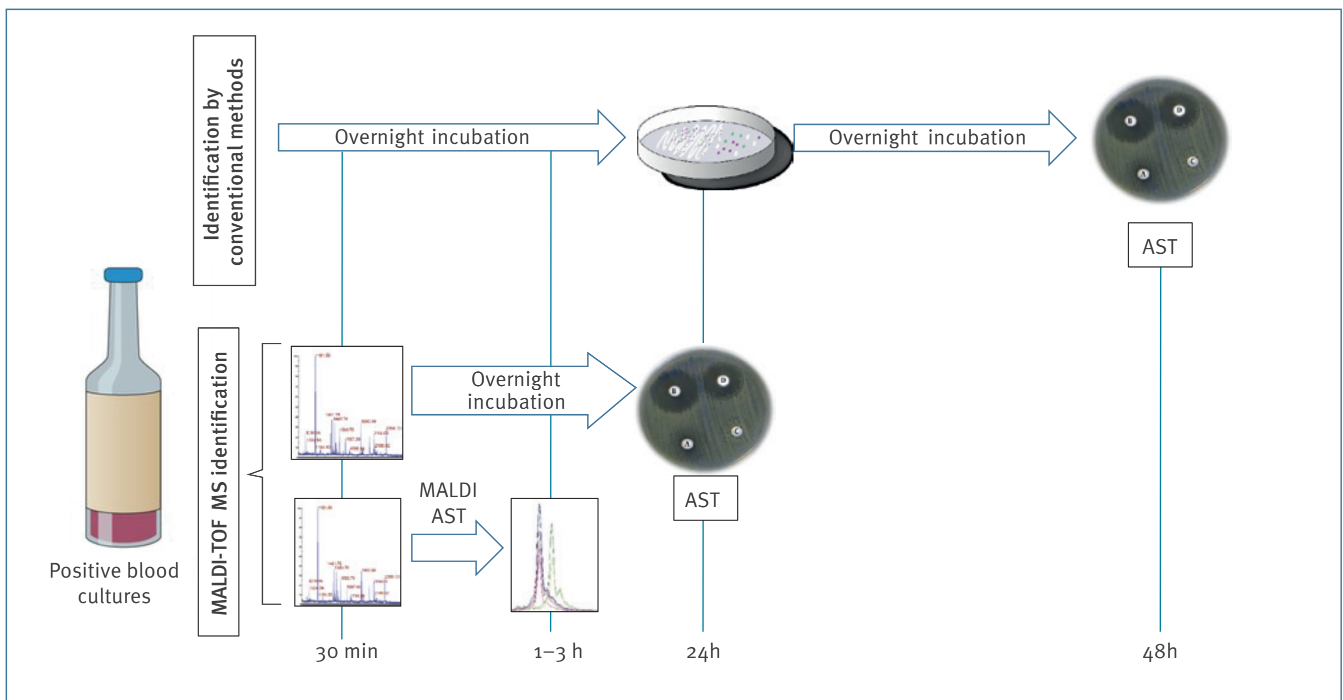
The rapid acquisition of protein spectra using MALDI-TOF MS has been implemented as a diagnostic tool for the identification of infection markers. For this purpose, the spectra are usually obtained directly from clinical samples, mainly serum or whole blood [98-100]. This approach could be useful in instances where the pathogen is seldom detected, as is often the case for suspected but unconfirmed fungal infections and for slow-growing microorganisms such as some *Mycobacterium* species. Precisely for these two applications, several authors have recently published interesting data [99-102].

Biomarkers for diagnosing fungal pathogens

In the case of fungal infections, Sendid et al. published the first evidence of the presence in serum samples of a disaccharide directly related to experimental invasive candidiasis in a mouse model and also in human sera. They further simplified this methodology and implemented it as routine identification of this biomarker from serum of patients with invasive candidiasis, invasive aspergillosis and mucormycoses [100]. Their results showed that the detection of the disaccharide marker (365 m/z) performed similarly to beta-D-glucan

Figure 2

Workflow for the identification of microorganisms from positive blood, review of MALDI-TOF MS use in public health and hospital hygiene, 2018



AST: antimicrobial susceptibility testing.

On the upper lane, the identification is made from colonies grown on agar plates after a 24-hour incubation; after another 24 hours, AST is performed. In the central lane, the microorganism is identified after 30 minutes using MALDI-TOF MS, but AST is delayed 24 hours. On the bottom lane, optimised MALDI-TOF MS identification and AST is performed within 3.5 hours, completing both tests within one working shift.

and galactomannan, thus complementing those tests. Although the detection of this biomarker has not been validated yet, its implementation could represent a rapid, inexpensive and easy-to-perform means for detecting invasive infections caused by a wide range of fungal species.

The detection of acute phase proteins with MALDI-TOF MS has also been tested as a marker of antifungal treatment response in a rabbit model of invasive pulmonary aspergillosis [103]. Although these proteins are not specific to fungal infection, their presence in infected rabbits was confirmed, as well as important changes in their expression as a response to antifungal treatment.

Biomarkers for diagnosing active and latent *Mycobacterium tuberculosis* infection

Few studies reported the identification of specific plasma biomarkers for latent tuberculosis infection (LTBI), using MALDI-TOF that could differentiate between healthy individuals and those with LTBI. In their study, Zhang et al. (2014) used weak cation exchange magnetic beads (MB-WCX Kit, Bruker Daltonics) to recover plasma proteins even in low concentration. They then acquired spectra of plasma proteins and analysed them with specific algorithms. This combination allowed them to develop a model to discriminate between healthy and LTBI individuals, based on the presence/absence of specific peaks [102]. The same concept was also developed by Sandhu et al., who detected three regions along the protein spectra (around 5.8kDa, 11.5kDa and 21kDa) of plasma samples that also allowed discrimination of healthy individuals from patients with active TB infection and symptomatic LTBI patients with 87–90% accuracy [101]. The advantage of these approaches is that the methodology can be easily standardised, thanks to the use of the commercial kit for protein recovery from plasma. However, the protein ranges analysed by both studies are different and so are the results obtained in both cases. The identification of accurate biomarkers for TB infection would make MALDI-TOF MS a valuable screening tool, though the marker peaks need further confirmation by molecular or serological methods.

Biomarkers for diagnosing viral infections

Finally, a similar approach has been applied recently for the identification of a panel of 10 respiratory viruses from infected cell cultures [104]. The authors utilised four commonly used cell lines to establish a background of protein peaks derived from the cell

culture and then found specific viral peaks using reference viral strains. The marker peaks were also robustly found in cell cultures infected with viruses from clinical samples. The authors found this methodology to be poorly discriminatory for closely related viruses. The same authors also reported the discrimination of three poliovirus serotypes using MALDI-TOF MS [105].

Discussion and conclusions

The implementation of MALDI-TOF MS has changed the way many microorganisms of clinical and public health interest are identified. Anaerobic bacterial species, yeasts, mycobacteria and an increasing number of moulds can be reliably identified using this technology. This fact is reflected in the amount of literature about this subject published during the past 10 years (Figure 2). Although only articles referenced in PubMed have been reviewed here, the large number of publications in this database reporting the use of MALDI-TOF MS to rapidly identify a wide range of microorganisms with public health relevance worldwide provides an up-to-date overview of the role of MALDI-TOF MS in this field.

Despite the successful results reported using MALDI-TOF MS and the wide range of scenarios where these findings could be applied, further studies are necessary to standardise the applied procedures and to confirm the reproducibility of the results. In a recent study, the methodology applied for typing was evaluated in different laboratories [106]. Technical and biological replicates were analysed in order to assay the reproducibility of the marker peaks detected in different populations of microorganisms. Their results displayed a reproducibility of technical and biological replicates ranging between 96.8–99.4% and 47.6–94.4%, respectively. Thus, the authors proposed the evaluated technology as a first-line screening tool in outbreak analysis and epidemiological studies. In addition, the use of classifier algorithms and linear support vector machine (SVM) allowed the correct classification of the isolates used for validation. The implementation of these bioinformatics tools, together with standardised procedures and the available software, will turn MALDI-TOF MS into a reliable reference methodology for typing isolates. Free software such as MALDIQuant [31] or proprietary software like FlexAnalysis and ClinProTools (Bruker Daltonics) or Bionumerics (Applied Maths, Sint-Martens-Latem, Belgium) allow automatic analysis of large amounts of protein spectra and facilitates the application of different classifiers for the correct identification of bacterial populations.

Additionally, available databases constructed by MALDI-TOF MS users can now be accessed online for the accurate identification of certain groups of microorganisms (https://microbenet.cdc.gov/) [37,107,108]. Taking into account the great impact of MALDI-TOF MS during the past 10 years, the knowledge that has been acquired during this time and the great flexibility of the technique, we think that its influence in public health will only become bigger in the coming years. Its use for resistance mechanism detection, typing and peak biomarker identification makes MALDI-TOF MS an excellent tool for monitoring the epidemiology of highly resistant or virulent pathogens, for outbreak detection and for screening of isolates within an outbreak, as the rapid acquisition and analysis of the protein spectra would facilitate prompt implementation of isolation measures and the identification of the affected patients. DNA sequencing tests could, therefore, be used as a confirmatory test only, to save time and resources.

Acknowledgements

BR-S (CPI14/00220) is a recipient of a Miguel Servet contract (ISCIII-MICINN) from Carlos III Health Institute (ISCIII), Madrid, Spain, partially financed by the by the European Regional Development Fund (FEDER) 'A way of making Europe', supported by the FIS program. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

The authors want to acknowledge Dr Francisco Parras for his expert advice concerning public health and critical review of this manuscript.

Conflict of interest

None declared.

Authors' contributions

BR-S and GG contributed to the article conception and organisation, and wrote and reviewed the manuscript. EC and ATC wrote and analysed the application of MALDI-TOF for AST, wrote this part and reviewed the final manuscript.

References

1. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36(2):380-407. https://doi.org/10.1111/j.1574-6976.2011.00298.x PMID: 22092265
2. Lévesque S, Dufresne PJ, Soualhine H, Domingo MC, Bekal S, Lefebvre B, et al. A Side by Side Comparison of Bruker Biotyper and VITEK MS: Utility of MALDI-TOF MS Technology for Microorganism Identification in a Public Health Reference Laboratory. PLoS One. 2015;10(12):e0144878. https://doi.org/10.1371/journal.pone.0144878 PMID: 26658918
3. Spanu T, Posteraro B, Fiori B, D'Inzeo T, Campoli S, Ruggeri A, et al. Direct maldi-tof mass spectrometry assay of blood culture broths for rapid identification of Candida species causing bloodstream infections: an observational study in two large microbiology laboratories. J Clin Microbiol. 2012;50(1):176-9. https://doi.org/10.1128/JCM.05742-11 PMID: 22090401
4. Martiny D, Debaugnies F, Gateff D, Gérard M, Aoun M, Martin C, et al. Impact of rapid microbial identification directly from positive blood cultures using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on patient management. Clin Microbiol Infect. 2013;19(12):E568-81. https://doi.org/10.1111/1469-0691.12282 PMID: 23890423
5. Clerc O, Prod'hom G, Vogne C, Bizzini A, Calandra T, Greub G. Impact of matrix-assisted laser desorption ionization time-of-flight mass spectrometry on the clinical management of patients with Gram-negative bacteremia: a prospective observational study. Clin Infect Dis. 2013;56(8):1101-7. https://doi.org/10.1093/cid/cis1204 PMID: 23264363
6. Oviaño M, Fernández B, Fernández A, Barba MJ, Mouriño C, Bou G. Rapid detection of enterobacteriaceae producing extended spectrum beta-lactamases directly from positive blood cultures by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Clin Microbiol Infect.

2014;20(11):1146-57. https://doi.org/10.1111/1469-0691.12729 PMID: 24942177

7. Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, et al. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant Staphylococcus aureus lineages. Int J Med Microbiol. 2011;301(1):64-8. https://doi.org/10.1016/j.ijmm.2010.06.002 PMID: 20728405
8. Freitas AR, Sousa C, Novais C, Silva L, Ramos H, Coque TM, et al. Rapid detection of high-risk Enterococcus faecium clones by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Diagn Microbiol Infect Dis. 2017;87(4):299-307. https://doi.org/10.1016/j.diagmicrobio.2016.12.007 PMID: 28109550
9. Christner M, Trusch M, Rohde H, Kwiatkowski M, Schlüter H, Wolters M, et al. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-Toxigenic Escherichia coli. PLoS One. 2014;9(7):e101924. https://doi.org/10.1371/journal.pone.0101924 PMID: 25003758
10. Pranada AB, Witt E, Bienia M, Kostrzewa M, Timke M. Accurate differentiation of Mycobacterium chimaera from Mycobacterium intracellulare by MALDI-TOF MS analysis. J Med Microbiol. 2017;66(5):670-7. https://doi.org/10.1099/jmm.0.000469 PMID: 28504926
11. Patel R. MALDI-TOF MS for the diagnosis of infectious diseases. Clin Chem. 2015;61(1):100-11. https://doi.org/10.1373/clinchem.2014.221770 PMID: 25278500
12. Wieser A, Schneider L, Jung J, Schubert S. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). Appl Microbiol Biotechnol. 2012;93(3):965-74. https://doi.org/10.1007/s00253-011-3783-4 PMID: 22198716
13. Stevenson LG, Drake SK, Murray PR. Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48(2):444-7. https://doi.org/10.1128/JCM.01541-09 PMID: 19955282
14. Prod'hom G, Bizzini A, Durussel C, Bille J, Greub G. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for direct bacterial identification from positive blood culture pellets. J Clin Microbiol. 2010;48(4):1481-3. https://doi.org/10.1128/JCM.01780-09 PMID: 20164269
15. Drancourt M. Detection of microorganisms in blood specimens using matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a review. Clin Microbiol Infect. 2010;16(11):1620-5. https://doi.org/10.1111/j.1469-0691.2010.03290.x PMID: 20545958
16. Opota O, Croxatto A, Prod'hom G, Greub G. Blood culture-based diagnosis of bacteraemia: state of the art. Clin Microbiol Infect. 2015;21(4):313-22. https://doi.org/10.1016/j.cmi.2015.01.003 PMID: 25753137
17. Rodríguez-Sánchez B, Sánchez-Carrillo C, Ruiz A, Marín M, Cercenado E, Rodríguez-Créixems M, et al. Direct identification of pathogens from positive blood cultures using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Clin Microbiol Infect. 2014;20(7):O421-7. https://doi.org/10.1111/1469-0691.12455 PMID: 24237623
18. Christner M, Rohde H, Wolters M, Sobottka I, Wegscheider K, Aepfelbacher M. Rapid identification of bacteria from positive blood culture bottles by use of matrix-assisted laser desorption-ionization time of flight mass spectrometry fingerprinting. J Clin Microbiol. 2010;48(5):1584-91. https://doi.org/10.1128/JCM.01831-09 PMID: 20237093
19. Schieffer KM, Tan KE, Stamper PD, Somogyi A, Andrea SB, Wakefield T, et al. Multicenter evaluation of the Sepsityper™ extraction kit and MALDI-TOF MS for direct identification of positive blood culture isolates using the BD BACTEC™ FX and VersaTREK(®) diagnostic blood culture systems. J Appl Microbiol. 2014;116(4):934-41. https://doi.org/10.1111/jam.12434 PMID: 24410849
20. Yan Y, He Y, Maier T, Quinn C, Shi G, Li H, et al. Improved identification of yeast species directly from positive blood culture media by combining Sepsityper specimen processing and Microflex analysis with the matrix-assisted laser desorption ionization Biotyper system. J Clin Microbiol. 2011;49(7):2528-32. https://doi.org/10.1128/JCM.00339-11 PMID: 21543564
21. Bidart M, Bonnet I, Hennebique A, Kherraf ZE, Pelloux H, Berger F, et al. An in-house assay is superior to Sepsityper for direct matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry identification of yeast species in blood cultures. J Clin Microbiol. 2015;53(5):1761-4. https://doi.org/10.1128/JCM.03600-14 PMID: 25762771
22. Leli C, Cenci E, Cardaccia A, Moretti A, D'Alò F, Pagliochini R, et al. Rapid identification of bacterial and fungal pathogens from positive blood cultures by MALDI-TOF MS. Int J Med Microbiol. 2013;303(4):205-9. https://doi.org/10.1016/j.ijmm.2013.03.002 PMID: 23602511
23. Croxatto A, Prod'hom G, Durussel C, Greub G. Preparation of a blood culture pellet for rapid bacterial identification and antibiotic susceptibility testing. J Vis Exp. 2014;(92):e51985. PMID:25350577
24. Idelevich EA, Schüle I, Grünastel B, Wüllenweber J, Peters G, Becker K. Rapid identification of microorganisms from positive blood cultures by MALDI-TOF mass spectrometry subsequent to very short-term incubation on solid medium. Clin Microbiol Infect. 2014;20(10):1001-6. https://doi.org/10.1111/1469-0691.12640 PMID: 24698361
25. Hoyos-Mallecot Y, Riazzo C, Miranda-Casas C, Rojo-Martín MD, Gutiérrez-Fernández J, Navarro-Marí JM. Rapid detection and identification of strains carrying carbapenemases directly from positive blood cultures using MALDI-TOF MS. J Microbiol Methods. 2014;105:98-101. https://doi.org/10.1016/j.mimet.2014.07.016 PMID: 25063679
26. Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF-mass spectrometry applications in clinical microbiology. Future Microbiol. 2010;5(11):1733-54. https://doi.org/10.2217/fmb.10.127 PMID: 21133692
27. Perez KK, Olsen RJ, Musick WL, Cernoch PL, Davis JR, Land GA, et al. Integrating rapid pathogen identification and antimicrobial stewardship significantly decreases hospital costs. Arch Pathol Lab Med. 2013;137(9):1247-54. https://doi.org/10.5858/arpa.2012-0651-OA PMID: 23216247
28. Osthoff M, Gürtler N, Bassetti S, Balestra G, Marsch S, Pargger H, et al. Impact of MALDI-TOF-MS-based identification directly from positive blood cultures on patient management: a controlled clinical trial. Clin Microbiol Infect. 2017;23(2):78-85. https://doi.org/10.1016/j.cmi.2016.08.009 PMID: 27569710
29. Dieckmann R, Malorny B. Rapid screening of epidemiologically important Salmonella enterica subsp. enterica serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol. 2011;77(12):4136-46. https://doi.org/10.1128/AEM.02418-10 PMID: 21515723
30. Ojima-Kato T, Yamamoto N, Nagai S, Shima K, Akiyama Y, Ota J, et al. Application of proteotyping Strain Solution™ ver. 2 software and theoretically calculated mass database in MALDI-TOF MS typing of Salmonella serotype. Appl Microbiol Biotechnol. 2017;101(23-24):8557-69. https://doi.org/10.1007/s00253-017-8563-3 PMID: 29032472
31. Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics. 2012;28(17):2270-1. https://doi.org/10.1093/bioinformatics/bts447 PMID: 22796955
32. Gasanov U, Hughes D, Hansbro PM. Methods for the isolation and identification of Listeria spp. and Listeria monocytogenes: a review. FEMS Microbiol Rev. 2005;29(5):851-75. https://doi.org/10.1016/j.femsre.2004.12.002 PMID: 16219509
33. Jadhav S, Gulati V, Fox EM, Karpe A, Beale DJ, Sevior D, et al. Rapid identification and source-tracking of Listeria monocytogenes using MALDI-TOF mass spectrometry. Int J Food Microbiol. 2015;202:1-9. https://doi.org/10.1016/j.ijfoodmicro.2015.01.023 PMID: 25747262
34. Hsueh PR, Lee TF, Du SH, Teng SH, Liao CH, Sheng WH, et al. Bruker biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of Nocardia, Rhodococcus, Kocuria, Gordonia, Tsukamurella, and Listeria species. J Clin Microbiol. 2014;52(7):2371-9. https://doi.org/10.1128/JCM.00456-14 PMID: 24759706
35. Bruker Daltonic Gmb H. MALDI Biotyper Subtyping Module. Bremen: Bruker Daltonic GmbH. [Accessed4 Jan 2019]. Available from: https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/Brochures/1851663_MBT_Subtyping_brochure_04-2017_ebook.pdf
36. Rizzardi K, Åkerlund T. High Molecular Weight Typing with MALDI-TOF MS - A Novel Method for Rapid Typing of Clostridium difficile. PLoS One. 2015;10(4):e0122457. https://doi.org/10.1371/journal.pone.0122457 PMID: 25923527
37. Graells T, Hernández-García M, Pérez-Jové J, Guy L, Padilla E. Legionella pneumophila recurrently isolated in a Spanish hospital: Two years of antimicrobial resistance surveillance. Environ Res. 2018;166:638-46. https://doi.org/10.1016/j.envres.2018.06.045 PMID: 29982152
38. Trnková K, Kotrbancová M, Špaleková M, Fulová M, Boledovičová J, Vesteg M. MALDI-TOF MS analysis as a useful tool for an identification of Legionella pneumophila, a facultatively pathogenic bacterium interacting with free-living amoebae: A case study from water supply system of hospitals in Bratislava (Slovakia). Exp Parasitol. 2018;184:97-102. https://doi.org/10.1016/j.exppara.2017.12.002 PMID: 29225047

39. Marín M, Cercenado E, Sánchez-Carrillo C, Ruiz A, Gómez González Á, Rodríguez-Sánchez B, et al. Accurate Differentiation of Streptococcus pneumoniae from other Species within the Streptococcus mitis Group by Peak Analysis Using MALDI-TOF MS. Front Microbiol. 2017;8:698. https://doi.org/10.3389/fmicb.2017.00698 PMID: 28487677
40. Harju I, Lange C, Kostrzewa M, Maier T, Rantakokko-Jalava K, Haanperä M. Improved Differentiation of Streptococcus pneumoniae and Other S. mitis Group Streptococci by MALDI Biotyper Using an Improved MALDI Biotyper Database Content and a Novel Result Interpretation Algorithm. J Clin Microbiol. 2017;55(3):914-22. https://doi.org/10.1128/JCM.01990-16 PMID: 28053215
41. World Health Organization (WHO). Global Tuberculosis Report 2017. Geneva: WHO; 2017. Available from: http://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf
42. Jagielski T, Minias A, van Ingen J, Rastogi N, Brzostek A, Żaczek A, et al. Methodological and Clinical Aspects of the Molecular Epidemiology of Mycobacterium tuberculosis and Other Mycobacteria. Clin Microbiol Rev. 2016;29(2):239-90. https://doi.org/10.1128/CMR.00055-15 PMID: 26912567
43. Rodríguez-Sánchez B, Ruiz-Serrano MJ, Ruiz A, Timke M, Kostrzewa M, Bouza E. Evaluation of MALDI Biotyper Mycobacteria Library v3.0 for Identification of Nontuberculous Mycobacteria. J Clin Microbiol. 2016;54(4):1144-7. https://doi.org/10.1128/JCM.02760-15 PMID: 26842704
44. Alcaide F, Amlerová J, Bou G, Ceyssens PJ, Coll P, Corcoran D, et al. How to: identify non-tuberculous Mycobacterium species using MALDI-TOF mass spectrometry. Clin Microbiol Infect. 2018;24(6):599-603. https://doi.org/10.1016/j.cmi.2017.11.012 PMID: 29174730
45. van Eck K, Faro D, Wattenberg M, de Jong A, Kuipers S, van Ingen J. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Fails To Identify Nontuberculous Mycobacteria from Primary Cultures of Respiratory Samples. J Clin Microbiol. 2016;54(7):1915-7. https://doi.org/10.1128/JCM.00304-16 PMID: 27147723
46. Kehrmann J, Schoerding AK, Murali R, Wessel S, Koehling HL, Mosel F, et al. Performance of Vitek MS in identifying nontuberculous mycobacteria from MGIT liquid medium and Lowenstein-Jensen solid medium. Diagn Microbiol Infect Dis. 2016;84(1):43-7. https://doi.org/10.1016/j.diagmicrobio.2015.10.007 PMID: 26527059
47. Fangous MS, Mougari F, Gouriou S, Calvez E, Raskine L, Cambau E, et al. Classification algorithm for subspecies identification within the Mycobacterium abscessus species, based on matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2014;52(9):3362-9. https://doi.org/10.1128/JCM.00788-14 PMID: 25009048
48. Kehrmann J, Wessel S, Murali R, Hampel A, Bange FC, Buer J, et al. Principal component analysis of MALDI TOF MS mass spectra separates M. abscessus (sensu stricto) from M. massiliense isolates. BMC Microbiol. 2016;16(1):24. https://doi.org/10.1186/s12866-016-0636-4 PMID: 26926762
49. Svensson E, Jensen ET, Rasmussen EM, Folkvardsen DB, Norman A, Lillebaek T. Mycobacterium chimaera in Heater-Cooler Units in Denmark Related to Isolates from the United States and United Kingdom. Emerg Infect Dis. 2017;23(3):507-9. https://doi.org/10.3201/eid2303.161941 PMID: 28035898
50. Lasch P, Wahab T, Weil S, Pályi B, Tomaso H, Zange S, et al. Identification of Highly Pathogenic Microorganisms by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: Results of an Interlaboratory Ring Trial. J Clin Microbiol. 2015;53(8):2632-40. https://doi.org/10.1128/JCM.00813-15 PMID: 26063856
51. Tracz DM, Antonation KS, Corbett CR. Verification of a Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Method for Diagnostic Identification of High-Consequence Bacterial Pathogens. J Clin Microbiol. 2016;54(3):764-7. https://doi.org/10.1128/JCM.02709-15 PMID: 26677252
52. Rudrik JT, Soehnlen MK, Perry MJ, Sullivan MM, Reiter-Kintz W, Lee PA, et al. Safety and Accuracy of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Identification of Highly Pathogenic Organisms. J Clin Microbiol. 2017;55(12):3513-29. https://doi.org/10.1128/JCM.01023-17 PMID: 29021156
53. Hu YY, Cai JC, Zhou HW, Zhang R, Chen GX. Rapid detection of porins by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Front Microbiol. 2015;6:784. https://doi.org/10.3389/fmicb.2015.00784 PMID: 26300858
54. Griffin PM, Price GR, Schooneveldt JM, Schlebusch S, Tilse MH, Urbanski T, et al. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. J Clin Microbiol. 2012;50(9):2918-31. https://doi.org/10.1128/JCM.01000-12 PMID: 22740710
55. Camara JE, Hays FA. Discrimination between wild-type and ampicillin-resistant Escherichia coli by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Bioanal Chem. 2007;389(5):1633-8. https://doi.org/10.1007/s00216-007-1558-7 PMID: 17849103
56. Mirande C, Canard I, Buffet Croix Blanche S, Charrier JP, van Belkum A, Welker M, et al. Rapid detection of carbapenemase activity: benefits and weaknesses of MALDI-TOF MS. Eur J Clin Microbiol Infect Dis. 2015;34(11):2225-34. https://doi.org/10.1007/s10096-015-2473-z PMID: 26337432
57. Sparbier K, Schubert S, Kostrzewa M. MBT-ASTRA: A suitable tool for fast antibiotic susceptibility testing? Methods. 2016;104:48-54. https://doi.org/10.1016/j.ymeth.2016.01.008 PMID: 26804565
58. Jung JS, Hamacher C, Gross B, Sparbier K, Lange C, Kostrzewa M, et al. Evaluation of a Semiquantitative Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Method for Rapid Antimicrobial Susceptibility Testing of Positive Blood Cultures. J Clin Microbiol. 2016;54(11):2820-4. https://doi.org/10.1128/JCM.01131-16 PMID: 27629893
59. Ceyssens PJ, Soetaert K, Timke M, Van den Bossche A, Sparbier K, De Cremer K, et al. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Combined Species Identification and Drug Sensitivity Testing in Mycobacteria. J Clin Microbiol. 2017;55(2):624-34. https://doi.org/10.1128/JCM.02089-16 PMID: 28003422
60. Edwards-Jones V, Claydon MA, Evason DJ, Walker J, Fox AJ, Gordon DB. Rapid discrimination between methicillin-sensitive and methicillin-resistant Staphylococcus aureus by intact cell mass spectrometry. J Med Microbiol. 2000;49(3):295-300. https://doi.org/10.1099/0022-1317-49-3-295 PMID: 10707951
61. Du Z, Yang R, Guo Z, Song Y, Wang J. Identification of Staphylococcus aureus and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem. 2002;74(21):5487-91. https://doi.org/10.1021/ac020109k PMID: 12433077
62. Walker J, Fox AJ, Edwards-Jones V, Gordon DB. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant Staphylococcus aureus: media effects and inter-laboratory reproducibility. J Microbiol Methods. 2002;48(2-3):117-26. https://doi.org/10.1016/S0167-7012(01)00316-5 PMID: 11777562
63. Bernardo K, Pakulat N, Macht M, Krut O, Seifert H, Fleer S, et al. Identification and discrimination of Staphylococcus aureus strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Proteomics. 2002;2(6):747-53. https://doi.org/10.1002/1615-9861(200206)2:6‹747::AID-PROT747›3.0.CO;2-V PMID: 12112858
64. Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P. The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant Staphylococcus aureus. FEMS Microbiol Lett. 2006;255(2):233-9. https://doi.org/10.1111/j.1574-6968.2005.00060.x PMID: 16448500
65. Mather CA, Werth BJ, Sivagnanam S, SenGupta DJ, Butler-Wu SM. Rapid Detection of Vancomycin-Intermediate Staphylococcus aureus by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol. 2016;54(4):883-90. https://doi.org/10.1128/JCM.02428-15 PMID: 26763961
66. Asakura K, Azechi T, Sasano H, Matsui H, Hanaki H, Miyazaki M, et al. Rapid and easy detection of low-level resistance to vancomycin in methicillin-resistant Staphylococcus aureus by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. PLoS One. 2018;13(3):e0194212. https://doi.org/10.1371/journal.pone.0194212 PMID: 29522576
67. Nagy E, Becker S, Sóki J, Urbán E, Kostrzewa M. Differentiation of division I (cfiA-negative) and division II (cfiA-positive) Bacteroides fragilis strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Med Microbiol. 2011;60(11):1584-90. https://doi.org/10.1099/jmm.0.031336-0 PMID: 21680764
68. Rhoads DD, Wang H, Karichu J, Richter SS. The presence of a single MALDI-TOF mass spectral peak predicts methicillin resistance in staphylococci. Diagn Microbiol Infect Dis. 2016;86(3):257-61. https://doi.org/10.1016/j.diagmicrobio.2016.08.001 PMID: 27568365
69. Oviaño M, Gómara M, Barba MJ, Sparbier K, Pascual Á, Bou G. Quantitative and automated MALDI-TOF MS-based detection of the plasmid-mediated quinolone resistance determinant AAC(6)-Ib-cr in Enterobacteriaceae. J Antimicrob Chemother. 2017;72(10):2952-4. https://doi.org/10.1093/jac/dkx218 PMID: 29091187
70. Sparbier K, Schubert S, Weller U, Boogen C, Kostrzewa M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection

of resistance against β-lactam antibiotics. J Clin Microbiol. 2012;50(3):927-37. https://doi.org/10.1128/JCM.05737-11 PMID: 22205812
71. Wright GD. Bacterial resistance to antibiotics: enzymatic degradation and modification. Adv Drug Deliv Rev. 2005;57(10):1451-70. https://doi.org/10.1016/j.addr.2005.04.002 PMID: 15950313
72. Jung JS, Popp C, Sparbier K, Lange C, Kostrzewa M, Schubert S. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid detection of β-lactam resistance in Enterobacteriaceae derived from blood cultures. J Clin Microbiol. 2014;52(3):924-30. https://doi.org/10.1128/JCM.02691-13 PMID: 24403301
73. De Carolis E, Paoletti S, Nagel D, Vella A, Mello E, Palucci I, et al. A rapid diagnostic workflow for cefotaxime-resistant Escherichia coli and Klebsiella pneumoniae detection from blood cultures by MALDI-TOF mass spectrometry. PLoS One. 2017;12(10):e0185935. https://doi.org/10.1371/journal.pone.0185935 PMID: 28982134
74. Vogne C, Prod'hom G, Jaton K, Decosterd LA, Greub G. A simple, robust and rapid approach to detect carbapenemases in Gram-negative isolates by MALDI-TOF mass spectrometry: validation with triple quadripole tandem mass spectrometry, microarray and PCR. Clin Microbiol Infect. 2014;20(12):O1106-12. https://doi.org/10.1111/1469-0691.12715 PMID: 24930405
75. Oviaño M, Barba MJ, Fernández B, Ortega A, Aracil B, Oteo J, et al. Rapid Detection of OXA-48-Producing Enterobacteriaceae by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol. 2016;54(3):754-9. https://doi.org/10.1128/JCM.02496-15 PMID: 26677247
76. Burckhardt I, Zimmermann S. Using matrix-assisted laser desorption ionization-time of flight mass spectrometry to detect carbapenem resistance within 1 to 2.5 hours. J Clin Microbiol. 2011;49(9):3321-4. https://doi.org/10.1128/JCM.00287-11 PMID: 21795515
77. Studentova V, Papagiannitsis CC, Izdebski R, Pfeifer Y, Chudackova E, Bergerova T, et al. Detection of OXA-48-type carbapenemase-producing Enterobacteriaceae in diagnostic laboratories can be enhanced by addition of bicarbonates to cultivation media or reaction buffers. Folia Microbiol (Praha). 2015;60(2):119-29. https://doi.org/10.1007/s12223-014-0349-8 PMID: 25261959
78. Monteferrante CG, Sultan S, Ten Kate MT, Dekker LJ, Sparbier K, Peer M, et al. Evaluation of different pretreatment protocols to detect accurately clinical carbapenemase-producing Enterobacteriaceae by MALDI-TOF. J Antimicrob Chemother. 2016;71(10):2856-67. https://doi.org/10.1093/jac/dkw208 PMID: 27287232
79. Wang L, Han C, Sui W, Wang M, Lu X. MALDI-TOF MS applied to indirect carbapenemase detection: a validated procedure to clearly distinguish between carbapenemase-positive and carbapenemase-negative bacterial strains. Anal Bioanal Chem. 2013;405(15):5259-66. https://doi.org/10.1007/s00216-013-6913-2 PMID: 23584712
80. Hrabák J, Walková R, Studentová V, Chudácková E, Bergerová T. Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2011;49(9):3222-7. https://doi.org/10.1128/JCM.00984-11 PMID: 21775535
81. Hrabák J, Chudáčková E, Papagiannitsis CC. Detection of carbapenemases in Enterobacteriaceae: a challenge for diagnostic microbiological laboratories. Clin Microbiol Infect. 2014;20(9):839-53. https://doi.org/10.1111/1469-0691.12678 PMID: 24813781
82. Papagiannitsis CC, Študentová V, Izdebski R, Oikonomou O, Pfeifer Y, Petinaki E, et al. Matrix-assisted laser desorption ionization-time of flight mass spectrometry meropenem hydrolysis assay with NH4HCO3, a reliable tool for direct detection of carbapenemase activity. J Clin Microbiol. 2015;53(5):1731-5. https://doi.org/10.1128/JCM.03094-14 PMID: 25694522
83. Fajardo A, Hernando-Amado S, Oliver A, Ball G, Filloux A, Martinez JL. Characterization of a novel Zn2+-dependent intrinsic imipenemase from Pseudomonas aeruginosa. J Antimicrob Chemother. 2014;69(11):2972-8. https://doi.org/10.1093/jac/dku267 PMID: 25185138
84. Rotova V, Papagiannitsis CC, Skalova A, Chudejova K, Hrabak J. Comparison of imipenem and meropenem antibiotics for the MALDI-TOF MS detection of carbapenemase activity. J Microbiol Methods. 2017;137:30-3. https://doi.org/10.1016/j.mimet.2017.04.003 PMID: 28390706
85. Lee AWT, Lam JKS, Lam RKW, Ng WH, Lee ENL, Lee VTY, et al. Comprehensive Evaluation of the MBT STAR-BL Module for Simultaneous Bacterial Identification and β-Lactamase-Mediated Resistance Detection in Gram-Negative Rods from Cultured Isolates and Positive Blood Cultures. Front Microbiol. 2018;9:334. https://doi.org/10.3389/fmicb.2018.00334 PMID: 29527202
86. Dortet L, Tandé D, de Briel D, Bernabeu S, Lasserre C, Gregorowicz G, et al. MALDI-TOF for the rapid detection of carbapenemase-producing Enterobacteriaceae: comparison of the commercialized MBT STAR®-Carba IVD Kit with two in-house MALDI-TOF techniques and the RAPIDEC® CARBA NP. J Antimicrob Chemother. 2018;73(9):2352-9. https://doi.org/10.1093/jac/dky209 PMID: 29897463
87. Marinach C, Alanio A, Palous M, Kwasek S, Fekkar A, Brossas JY, et al. MALDI-TOF MS-based drug susceptibility testing of pathogens: the example of Candida albicans and fluconazole. Proteomics. 2009;9(20):4627-31. https://doi.org/10.1002/pmic.200900152 PMID: 19750514
88. De Carolis E, Vella A, Florio AR, Posteraro P, Perlin DS, Sanguinetti M, et al. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry for caspofungin susceptibility testing of Candida and Aspergillus species. J Clin Microbiol. 2012;50(7):2479-83. https://doi.org/10.1128/JCM.00224-12 PMID: 22535984
89. Vella A, De Carolis E, Vaccaro L, Posteraro P, Perlin DS, Kostrzewa M, et al. Rapid antifungal susceptibility testing by matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis. J Clin Microbiol. 2013;51(9):2964-9. https://doi.org/10.1128/JCM.00903-13 PMID: 23824764
90. Saracli MA, Fothergill AW, Sutton DA, Wiederhold NP. Detection of triazole resistance among Candida species by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). Med Mycol. 2015;53(7):736-42. https://doi.org/10.1093/mmy/myv046 PMID: 26162474
91. Stupar P, Opota O, Longo G, Prod'hom G, Dietler G, Greub G, et al. Nanomechanical sensor applied to blood culture pellets: a fast approach to determine the antibiotic susceptibility against agents of bloodstream infections. Clin Microbiol Infect. 2017;23(6):400-5. https://doi.org/10.1016/j.cmi.2016.12.028 PMID: 28062319
92. CLSI. CLSI document M100. Performance Standards for Antimicrobial Susceptibility Testing. Clinical and Laboratory Standards Institute, Wayne, PA. 2018
93. Lange C, Schubert S, Jung J, Kostrzewa M, Sparbier K. Quantitative matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid resistance detection. J Clin Microbiol. 2014;52(12):4155-62. https://doi.org/10.1128/JCM.01872-14 PMID: 25232164
94. Sauget M, Bertrand X, Hocquet D. Rapid antibiotic susceptibility testing on blood cultures using MALDI-TOF MS. PLoS One. 2018;13(10):e0205603. https://doi.org/10.1371/journal.pone.0205603 PMID: 30308072
95. Van Driessche L, Bokma J, Gille L, Ceyssens PJ, Sparbier K, Haesebrouck F, et al. Rapid detection of tetracycline resistance in bovine Pasteurella multocida isolates by MALDI Biotyper antibiotic susceptibility test rapid assay (MBT-ASTRA). Sci Rep. 2018;8(1):13599. https://doi.org/10.1038/s41598-018-31562-8 PMID: 30206239
96. Vatanshenassan M, Boekhout T, Lass-Flörl C, Lackner M, Schubert S, Kostrzewa M, et al. Proof of Concept for MBT ASTRA, a Rapid Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS)-Based Method To Detect Caspofungin Resistance in Candida albicans and Candida glabrata. J Clin Microbiol. 2018;56(9):e00420-18. https://doi.org/10.1128/JCM.00420-18 PMID: 30021820
97. Justesen US, Acar Z, Sydenham TV, Johansson ÅESGAI (ESCMID Study Group on Anaerobic Infections). Antimicrobial susceptibility testing of Bacteroides fragilis using the MALDI Biotyper antibiotic susceptibility test rapid assay (MBT-ASTRA). Anaerobe. 2018;54:236-9. https://doi.org/10.1016/j.anaerobe.2018.02.007 PMID: 29501419
98. Deng C, Lin M, Hu C, Li Y, Gao Y, Cheng X, et al. Establishing a serologic decision tree model of extrapulmonary tuberculosis by MALDI-TOF MS analysis. Diagn Microbiol Infect Dis. 2011;71(2):144-50. https://doi.org/10.1016/j.diagmicrobio.2011.06.021 PMID: 21855247
99. Sendid B, Poissy J, François N, Mery A, Courtecuisse S, Krzewinski F, et al. Preliminary evidence for a serum disaccharide signature of invasive Candida albicans infection detected by MALDI Mass Spectrometry. Clin Microbiol Infect. 2015;21(1):88.e1-6. https://doi.org/10.1016/j.cmi.2014.08.010 PMID: 25636941
100. Mery A, Sendid B, François N, Cornu M, Poissy J, Guerardel Y, et al. Application of Mass Spectrometry Technology to Early Diagnosis of Invasive Fungal Infections. J Clin Microbiol. 2016;54(11):2786-97. https://doi.org/10.1128/JCM.01655-16 PMID: 27605710
101. Sandhu G, Battaglia F, Ely BK, Athanasakis D, Montoya R, Valencia T, et al. Discriminating active from latent tuberculosis in patients presenting to community clinics.

PLoS One. 2012;7(5):e38080. https://doi.org/10.1371/journal.pone.0038080 PMID: 22666453

102. Zhang X, Liu F, Li Q, Jia H, Pan L, Xing A, et al. A proteomics approach to the identification of plasma biomarkers for latent tuberculosis infection. Diagn Microbiol Infect Dis. 2014;79(4):432-7. https://doi.org/10.1016/j.diagmicrobio.2014.04.005 PMID: 24865408

103. Krel M, Petraitis V, Petraitiene R, Jain MR, Zhao Y, Li H, et al. Host biomarkers of invasive pulmonary aspergillosis to monitor therapeutic response. Antimicrob Agents Chemother. 2014;58(6):3373-8. https://doi.org/10.1128/AAC.02482-14 PMID: 24687510

104. Calderaro A, Arcangeletti MC, Rodighiero I, Buttrini M, Montecchini S, Vasile Simone R, et al. Identification of different respiratory viruses, after a cell culture step, by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). Sci Rep. 2016;6(1):36082. https://doi.org/10.1038/srep36082 PMID: 27786297

105. Calderaro A, Arcangeletti MC, Rodighiero I, Buttrini M, Gorrini C, Motta F, et al. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. Sci Rep. 2014;4(1):6803. https://doi.org/10.1038/srep06803 PMID: 25354905

106. Oberle M, Wohlwend N, Jonas D, Maurer FP, Jost G, Tschudin-Sutter S, et al. The Technical and Biological Reproducibility of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) Based Typing: Employment of Bioinformatics in a Multicenter Study. PLoS One. 2016;11(10):e0164260. https://doi.org/10.1371/journal.pone.0164260 PMID: 27798637

107. Erler R, Wichels A, Heinemeyer EA, Hauk G, Hippelein M, Reyes NT, et al. VibrioBase: A MALDI-TOF MS database for fast identification of Vibrio spp. that are potentially pathogenic in humans. Syst Appl Microbiol. 2015;38(1):16-25. https://doi.org/10.1016/j.syapm.2014.10.009 PMID: 25466918

108. Normand AC, Becker P, Gabriel F, Cassagne C, Accoceberry I, Gari-Toussaint M, et al. Validation of a New Web Application for Identification of Fungi by Use of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol. 2017;55(9):2661-70. https://doi.org/10.1128/JCM.00263-17 PMID: 28637907

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

Perspective

Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157:H7, England, 2013 to 2017

Claire Jenkins[1], Timothy J Dallman[1], Kathie A Grant[1]
1. National Infection Service, Public Health England, United Kingdom

Correspondence: Claire Jenkins (claire.jenkins@phe.gov.uk)

Citation style for this article:
Jenkins Claire, Dallman Timothy J, Grant Kathie A. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup O157:H7, England, 2013 to 2017. Euro Surveill. 2019;24(4):pii=1800346. https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800346

Article submitted on 28 Jun 2018 / accepted on 25 Oct 2018 / published on 24 Jan 2019

We aim to provide insight and guidance on the utility of whole genome sequencing (WGS) data for investigating food-borne outbreaks of Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 in England between 2013 and 2017. Analysis of WGS data delivered an unprecedented level of strain discrimination when compared with multilocus variable number tandem repeat analysis. The robustness of the WGS method ensured confidence in the microbiological identification of linked cases, even when epidemiological links were obscured. There was evidence that phylogeny derived from WGS data can be used to trace the geographical origin of an isolate. Further analysis of the phylogenetic data provided insight on the evolutionary context of emerging pathogenic strains. Publically available WGS data linked to the clinical, epidemiological and environmental context of the sequenced strain has improved trace back investigations during outbreaks. Expanding the use of WGS-based typing analysis globally will ensure the rapid implementation of interventions to protect public health, inform risk assessment and facilitate the management of national and international food-borne outbreaks of STEC O157:H7.

Background

In the 1980s, the emergence of Shiga toxin-producing *E. coli* (STEC) O157:H7, and the increasing number of outbreaks of gastrointestinal disease and haemolytic uraemic syndrome (HUS) associated with this serotype, stimulated the development of subtyping methods that provided a higher level of strain discrimination than serotyping [1,2]. Phage typing was adopted by Public Health England (PHE) and is still used today [3]. In the1990s and 2000s, Pulsed field gel electrophoresis (PFGE) and multilocus variable number tandem repeat (VNTR) analysis (MLVA) respectively, were used reactively in outbreaks that had already been identified by epidemiological links, although from 2012 prospective typing using MLVA was also in use [4,5]. PHE implemented whole genome sequencing (WGS) as the molecular typing method of choice for all isolates of STEC O157:H7 in June 2015 [6]. The aim of this perspective is to summarise the evaluation and share experiences on the utility of whole genome sequencing (WGS) data for investigating food-borne outbreaks of STEC O157:H7 in England and discuss the impact of this approach on informing risk assessment and risk management of this clinically important food-borne pathogen.

Preliminary evaluation studies and comparisons with multilocus variable number tandem repeat analysis

To assess the epidemiological relevance of genetic similarity between genomes a retrospective comparison was performed in 2014. Randomly selected isolates (n = 572) from the bacterial strain collection archive held at PHE were sequenced in order to assess the applicability of a common source single nucleotide polymorphism (SNP) variation threshold for outbreak detection, based on temporal and epidemiological linkages between isolates [6]. This study showed that, at the core genome level, isolates of STEC O157:H7 greater than five SNPs different were less likely to be part of the same temporally linked outbreak than those less than five SNPs different. During these preliminary investigations, previously unidentified clusters of isolates that fell within five SNPs of each other were detected; however, all but two of these clusters were too small to support meaningful epidemiological analysis. Following further epidemiological investigations, one of the two larger clusters was ultimately linked to consumption of contaminated salad leaves and the other was associated with exposure to animals at a national park [6]. Subsequent studies revealed that

Table 1

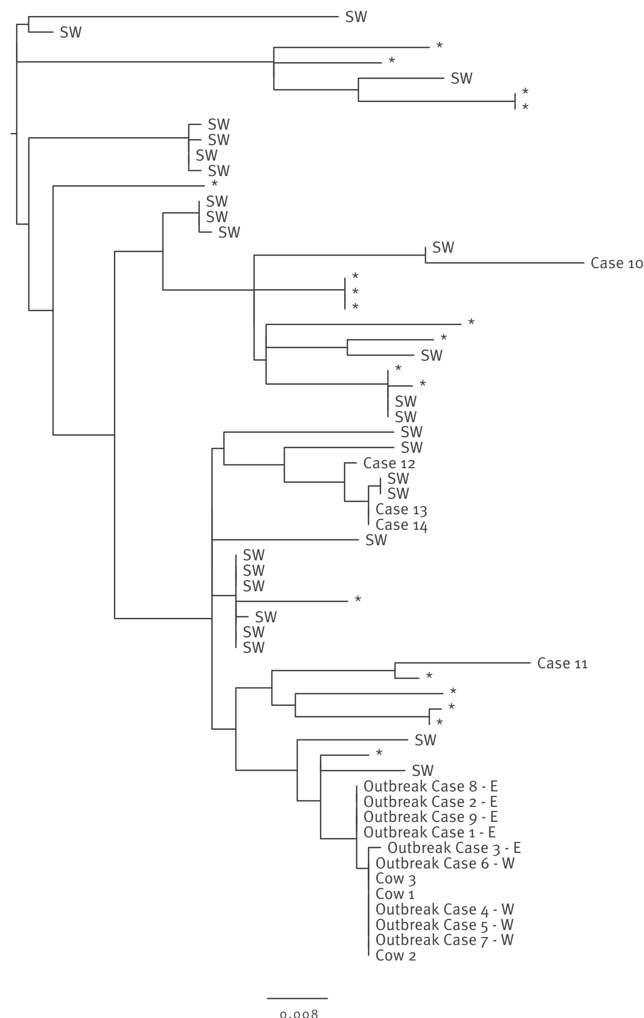
Use of whole genome sequencing to inform investigations of outbreaks of STEC O157:H7 and context, England 2013–2017

Reference	Description	Benefits of the WGS approach and context
Butcher et al. 2016 [7]	Outbreak of STEC O157:H7 PT21/28 associated with raw drinking milk in the south-west of England, 2014	• Robust, high level strain discrimination compared with traditional molecular typing methods • Forensic level typing for case ascertainment • Inferring the geographical origin of an outbreak strain from the phylogeny at the national level • Domestic source of outbreak strain • Evolutionary context of outbreak strains
Jenkins et al. 2015 [28]	Two national concurrent outbreaks of STEC O157:H7 PT2 associated with contaminated watercress, 2013	• Inferring the geographical origin of an outbreak strain from the phylogeny at the national and international level • Domestic source of outbreak strain • Non-domestic source of outbreak strain • Evolutionary context of outbreak strains
Mikhail et al. 2017 [8]	National outbreak of STEC O157:H7 PT8 associated with contaminated prepacked mixed leaf salad, 2015	• Inferring the geographical origin of an outbreak strain from the phylogeny at the national level • Domestic source of outbreak strain • Evolutionary context of outbreak strains
Byrne et al. 2016 [22]	Epidemiological and Microbiological Investigation of an Outbreak of Severe Disease from Shiga Toxin-Producing *Escherichia coli*O157 Infection Associated with Consumption of a Slaw Garnish	• Robust, high level strain discrimination compared with traditional molecular typing methods • Forensic level typing for case ascertainment • Domestic source of outbreak strain
Wilson et al. 2018 [24]	Outbreak of STEC O157:H7 PT21/28 associated with contaminated meat products at two butchers' premises in the north-east of England, 2015	• Forensic level typing for case ascertainment • Domestic source of outbreak strain
Rowell et al. 2016 [25]	Outbreak of STEC O157 PT21/28 associated with a lamb-feeding event	• Robust, high level strain discrimination compared with traditional molecular typing methods • Forensic level typing for case ascertainment • Domestic source of outbreak strain
Underwood et al. 2014 [27]	Outbreak of STEC O157:H7 at an open farm in the south-east of England, 2009	• Robust, high level strain discrimination compared with traditional molecular typing method • Domestic source of outbreak strain
Gobin et al. 2018 [31]	National outbreak of Shiga toxin producing *E. coli* O157: H7 linked to mixed salad leaves, 2016.	• Inferring the geographical origin of an outbreak strain from the phylogeny at the international level • Non-domestic source of outbreak strain • Evolutionary context of outbreak strains
Cowley et al. 2016 [32]	Two related sequential outbreaks of STEC O157:H7 PT8 and PT54 associated with the same restaurant, 2013	• Robust, high level strain discrimination compared with traditional molecular typing methods • Inferring the geographical origin of an outbreak strain from the phylogeny at the international level • Non-domestic source of outbreak strain • Evolutionary context of outbreak strains

STEC: Shiga toxin-producing *Escherichia coli*; WGS: whole genome sequencing.

Figure 1

Phylogenetic relationship between isolates from human Shiga toxin-producing *Escherichia coli* O157:H7 PT21/28 cases linked to consumption of raw milk and cattle, and isolates from sporadic human clinical cases that fell within a 25 SNP cluster of the outbreak isolates, England, 2014



RDM: raw drinking milk; MLVA: multilocus variable number tandem repeat analysis; SNP: single nucleotide polymorphism; WGS: whole genome sequencing.

Cases 1, 2, 3, 8 and 9, initially identified by epidemiological links (reporting the consumption of RDM from the same farm), are designated 'E'. Cases 4–7 initially identified by analysis of the WGS data (and subsequently found to have consumed RDM from the implicated farm), are designated 'W'. Cases 10–14 were identified as potentially linked to the outbreak by MLVA, but were shown not to be directly linked by WGS and subsequent epidemiological investigations. Cases designated SW resided in the south-west of England, but did not report consumption of RDM.

deeper phylogenetic relationships may provide epidemiologically useful information or associations [7,8] (Table 1).

Methods applied for sequencing and typing STEC O157:H7

DNA from isolates of STEC O157:H7 was extracted on the QiaSymphony (Qiagen, Germany), and sequenced on the HiSeq 2500 platform (Illumina Inc, United States (US)) yielding paired-end reads of 100bp in length. High quality reads were mapped to the reference STEC O157:H7 strain, Sakai (GenBank accession BA000007), using Burrows-Wheeler Aligner – Maximum Exact Matching (BWA MEM) [9]. The sequence alignment map output from BWA were sorted and indexed to produce a binary alignment map (BAM) using Samtools [9]. Genome Analysis Toolkit (GATK2) was then used to create a variant call format (VCF) file from each of the BAMs, which were further parsed to extract only SNP positions of high quality (mapping quality (MQ) › 30, depth (DP) › 10, variant ratio › 0.9) [10,11].

Hierarchical single linkage clustering was performed on the pairwise SNP difference between all isolates at descending distance thresholds (Δ250, Δ100, Δ50, Δ25, Δ10, Δ5, Δ0). The result of the clustering is a SNP profile, or SNP address, that is used to describe the population structure based on clonal group membership, as indicated by the number at each level of the seven-number SNP address [12]. Shiga toxin (Stx) subtyping was performed as described elsewhere [13].

Timely resolution and improved case ascertainment during outbreak investigations

Published studies comparing PFGE and MLVA, and WGS for typing STEC conclude that WGS is the superior technique [14-19]. Using a survival analysis, Dallman et al. [6] showed in a study published in 2015, that there was no significant temporal difference between MLVA and WGS SNP typing with respect to the time to identify a cluster, i.e. WGS was as sensitive as MLVA with respect to detecting an outbreak. However, when the time to cluster completion (the rate all cases of a cluster are clustered) from the initial cluster event (any two cases of a cluster are clustered) was reviewed, there was a significant speed increase in rate of completion of clusters with WGS when compared with MLVA. Other studies have also highlighted the considerable confidence WGS data affords in assigning 'like' vs 'not-like' status to two potentially linked bacteria [20].

This level of confidence in the microbiological typing data improves case ascertainment during outbreak investigations. In September 2014, the national enhanced STEC surveillance system [21] detected five cases associated with the consumption of raw drinking milk (RDM) produced at a farm in the south-west of England [7] (Table 1). Real time MLVA surveillance identified an additional nine isolates that appeared to be closely related to the outbreak strain by MLVA; there was uncertainty as to whether these additional cases were linked to the outbreak, as none of the cases reported RDM consumption on the STEC enhanced surveillance questionnaire (Figure 1, Cases 4–7, 10–14). Analysis of the WGS data revealed that four of these nine cases were part of the outbreak (Figure 1, Cases 4–7) and five were not associated with the outbreak (Figure 1, Cases 10–14). The nine cases were re-interviewed and asked

Table 2

Characteristics of domestic and non-domestic clusters of STEC O157:H7, England, 2013–2017

Domestic lineage, clade or cluster Sub-lineages Ic and IIb and clusters within sub-lineages IIc and I/II	Non-domestic lineage, clade Sub-lineages Ia, Ib and IIa and I/II
Characteristics	
Common in domestic dataset	Rare in domestic dataset
Short branch lengths (low level diversity) between clusters and clades representing frequent sampling of a restricted pool	Longer branch lengths (high level diversity) between clusters and clades representing infrequent sampling of the global pool
High frequency of domestic animal isolates sampled during prevalence studies and sequenced isolates are included in the dataset	Domestic animal isolates not present
Cases do not report recent travel outside the UK before onset of symptoms	High frequency of cases reporting recent travel before onset of symptoms
Cases from outbreaks known to be associated with domestically produced food,	Cases from outbreaks known to be associated with imported food, Outbreak Scenario 2
Cases associated with local environmental exposures, such as petting farms or parks	Cases not associated with local environmental exposures

STEC: Shiga toxin-producing *Escherichia coli*; UK: United Kingdom.

questions about their consumption of dairy products, a search was also carried out for their names and postcodes on the distribution list supplied by the operations manager at the implicated farm. Subsequent epidemiological investigations provided evidence that the four cases, identified by WGS as being linked to the outbreak, had consumed RDM but initially failed to recall an accurate food history or were unaware that the milk was unpasteurised; no evidence of consumption of RDM was uncovered for the remaining five cases identified by MLVA only [7]

It has been shown that epidemiological investigations are often confounded by poor patient recall of the food they consumed before onset of symptoms, particularly when the product is a side dish (e.g. salad leaves or raw vegetables) or an ingredient of the main dish (e.g. herbs or spices), so called 'stealth vehicles' [22,23]. The forensic-level microbiological typing provided by WGS can be used to generate a robust case definition for case ascertainment, even when the epidemiological links are obscured by poor patient recall of their history of food consumption [24,25].

Inferring the geographical origin and/or potential animal reservoir of a food-borne outbreak strain

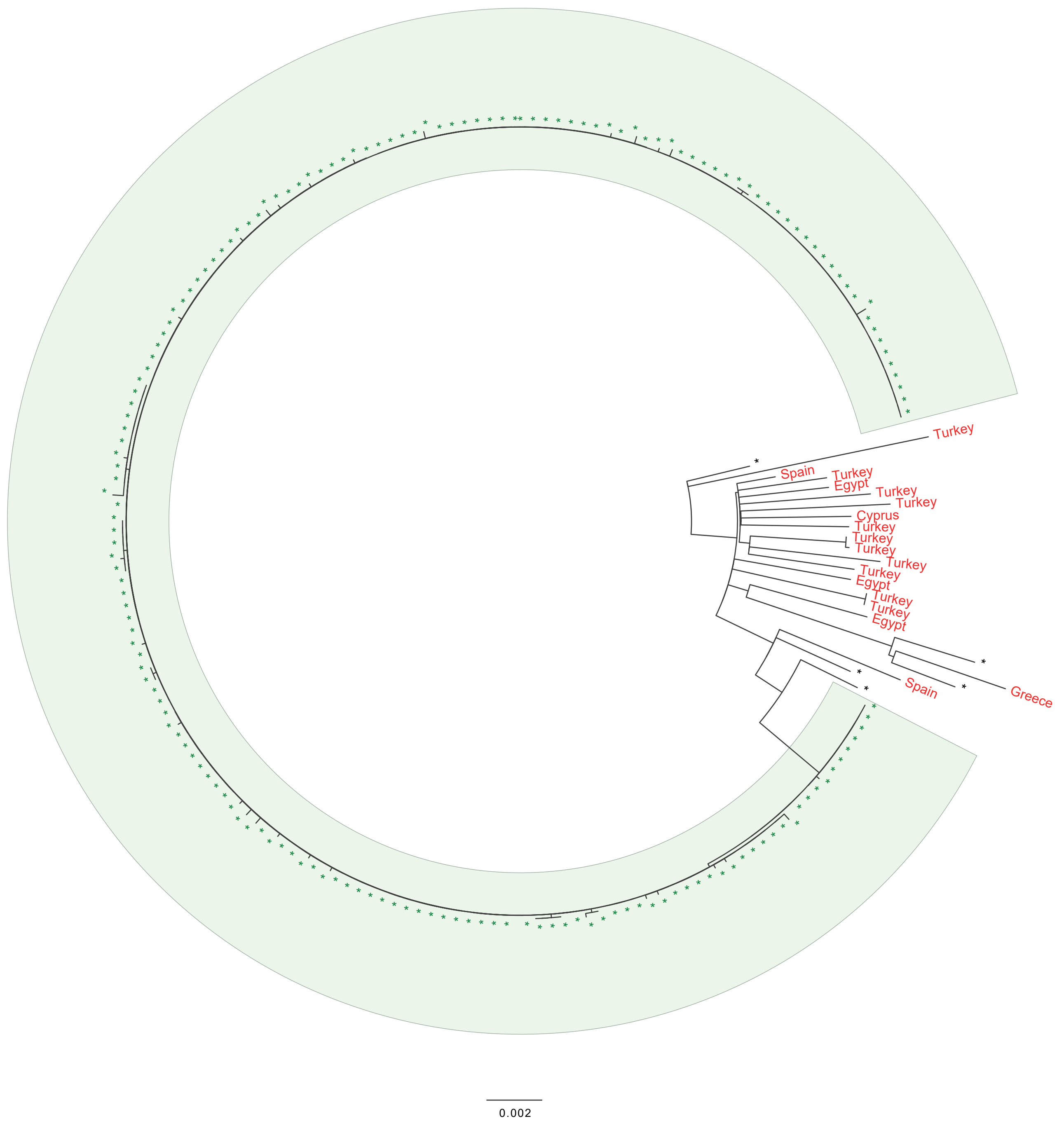
WGS data offers robust, high-level phylogenetic resolution and utilises quantifiable genetic markers that provide insight on the evolutionary context of an outbreak strain. Analysis of the data from the STEC O157:H7 dataset held at PHE showed that by exploring the context of the deeper phylogenetic relationship between isolates, the source of infection could be linked to specific geographical regions of the United Kingdom (UK). For example, the farm implicated in the RDM outbreak in 2014 was located in the south-west of England [7] (Table 1). Even though none of the cases within the same 25 SNP cluster as the outbreak strain reported consumption of RDM on the STEC-enhanced surveillance questionnaire, epidemiological analysis showed that 23/33 (70%) of these cases were resident in the south-west of England or had travelled there within 7 days before the onset of illness (Figure 1). Spatial analysis of the geographical location of the presumed exposure of the STEC O157:H7 cases within this 25 SNP cluster revealed a highly significant cluster in the south-west of England region. Rates of infection with this strain were significantly lower in other parts of England. This analysis provided evidence that the source of infection for outbreaks and sporadic cases of STEC O157:H7 in the UK may be geographically restricted and that it may be possible to map the location of the source using an phylogenetic approach, thus providing an evidence base to direct trace back investigations to specific locations.

Food-borne outbreaks – domestic or non-domestic origin?

Having investigated clusters within the UK that may be geographically restricted, the possibility that isolates of STEC O157:H7 may also exhibit geographical clustering on a global scale, was considered [26]. By tracking the expansion of the three major lineages, the sub-lineages and by superimposing epidemiological data onto the phylogeny e.g. known domestic exposures and recent travel abroad (less than 7 days before onset of symptoms), we can speculate that certain sub-lineages, clades, or clusters may be domestic or non-domestic. For example, sub-lineages Ic and IIb and certain clades and clusters in sub-lineages IIc and I/II are associated with UK strains, whereas sub-lineages 1a, 1b and IIa are likely to be imported from outside the UK. Strains belonging to domestic lineages were more common than non-domestic strains in the routine surveillance collection, and exhibited less diversity within clades because sampling of the restricted

Figure 2

Phylogenetic relationship between isolates associated with an outbreak of red Batavia salad leaves and those from resident cases reporting recent travel to countries in the Mediterranean region, United Kingdom, 2016



UK: United Kingdom.

Isolates associated with Batavia salad leaves are highlighted in green. Isolates from cases resident in the UK reporting recent travel to countries in the Mediterranean region are highlighted in red.

Quality trimmed Illumina reads were mapped to the STEC O157 reference genome Sakai (Genbank accession BA000007) using BWA-MEM. SNPs were identified using GATK2 in unified genotyper mode. Core genome positions that had a high quality SNP (>90% consensus, minimum depth 10x, MQ ≥ 30) in at least one isolate were extracted. SNP positions that were present in at least 80% of isolates were used to derive maximum likelihood phylogenies with RaxML using the GTRCAT model with 1,000 iterations.

pool of diversity (i.e. mainly UK cattle and sheep) was more frequent (Table 2). Included in the domestic clusters, were isolates from UK farm animals [7,24,25,27] and from cases associated with food-borne outbreaks where the food was identified as being of UK origin [7,22,24,28-30] (Table 1).

In contrast, non-domestic clades were more likely to be rare in the UK STEC O157:H7 surveillance database and associated with higher genetic diversity between isolates within a phylogenetic group; representing sparse sampling of a larger pool of diversity (i.e. a wide variety of zoonotic sources dispersed globally) (Table 2). Within these clades, no UK animal isolates were present, as the zoonotic source was located elsewhere, and a high proportion of isolates were from cases reporting foreign travel within 7 days of onset of symptoms (Table 2) [31]. Furthermore, the cases not reporting travel were linked to outbreaks associated with the consumption of imported herbs or salad leaves, or salad leaves grown in the UK from imported seed [28,31,32] (Table 1).

Analysis of WGS data from an outbreak in 2016, linked to the consumption of contaminated mixed leaf salad, revealed that the outbreak strain belonged to an uncommon clade in the PHE database and exhibited low levels of sampled diversity, characterised by longer branch lengths indicative of infrequent sampling from a widespread pool of strains [31]. The clade included a high proportion of cases reporting recent travel to Mediterranean countries, compared with other clades in the PHE database (Figure 2). Contaminated imported red Batavia lettuce leaves were suspected as the vehicle of infection, based on the exposure window assessment and supply chain timelines, although no microbiological evidence was obtained [31].

As more countries implement standardised, open access WGS data for routine surveillance of STEC, cross border exchange of WGS data will have a major impact on the ability to investigate national and international outbreaks of food-borne disease [33,34].

Conclusions

This perspective providing an overview of the use of WGS data during food-borne outbreak investigations in the United Kingdom demonstrated a number of advantages of using this approach: (i) unprecedented level of strain discrimination; (ii) robust, stable genetic markers; (iii) case identification when epidemiological links are obscured; (iv) geographical origins of outbreak strains may be inferred from the phylogenetic signal; and (v) insight into the evolutionary context for emerging pathogenic strains.

We found that collecting detailed epidemiological data is essential to best interpret phylogenetic clusters and that by defining clusters by the number of SNP differences between isolates provides information on strain relatedness. The central tenet of WGS based typing is that the fewer nucleotide differences between a pair of isolates the less time since divergence from a common ancestor i.e. isolates are more likely to originate from the same source population. The amount of diversity sampled when analysing a source population is dependent on the effective size of the population and the duration of infection. Therefore, it is not prudent to define absolute thresholds of nucleotide difference for inclusion and exclusion of isolates within an outbreak and epidemiological information should always be used, where possible, to inform the outbreak definition.

Expanding the use of WGS based typing analysis globally will improve trace back investigations in the event of a food-borne outbreak, ensuring the rapid implementation of interventions to protect public health. For the purposes of risk assessment and management of food-borne outbreaks, the utility of publicly available WGS database linked to the clinical, epidemiological and environmental context of each strain cannot be underestimated.

Acknowledgements

We would like to thank Francesco Tripodo, Vivienne do Nascimento, Amy Gentle, Marie Anne Chattaway, Dawn Hedges and Michael Wright at GBRU. We would also like to thank the national Gastrointestinal Infections national surveillance team, specifically Lisa Byrne, Natalie Adams, Amy Mikhail and Bob Adak. This work was supported by the National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections and the Department of Environment, Food and Rural Affairs. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Conflict of interest

None declared.

Authors' contributions

Claire Jenkins wrote the manuscript. Claire Jenkins, Timothy Dallman and Kathie Grant all read, commented upon and approved the final version of the paper.

References

1. Taylor CM, White RH, Winterborn MH, Rowe B. Haemolytic-uraemic syndrome: clinical experience of an outbreak in the West Midlands. Br Med J (Clin Res Ed). 1986;292(6534):1513-6. https://doi.org/10.1136/bmj.292.6534.1513 PMID: 3087499
2. Wall PG, McDonnell RJ, Adak GK, Cheasty T, Smith HR, Rowe B. General outbreaks of vero cytotoxin producing Escherichia coli O157 in England and Wales from 1992 to 1994. Commun Dis Rep CDR Rev. 1996;6(2):R26-33. PMID: 8777442
3. Khakhria R, Duck D, Lior H. Extended phage-typing scheme for Escherichia coli O157:H7. Epidemiol Infect. 1990;105(3):511-20. https://doi.org/10.1017/S0950268800048135 PMID: 2249715
4. Willshaw GA, Smith HR, Cheasty T, Wall PG, Rowe B. Vero cytotoxin-producing Escherichia coli O157 outbreaks in England and Wales, 1995: phenotypic methods and genotypic

subtyping. Emerg Infect Dis. 1997;3(4):561-5. https://doi.org/10.3201/eid0304.970422 PMID: 9366610
5. Byrne L, Elson R, Dallman TJ, Perry N, Ashton P, Wain J, et al. Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing Escherichia coli O157 as a routine public health tool in England. PLoS One. 2014;9(1):e85901. https://doi.org/10.1371/journal.pone.0085901 PMID: 24465775
6. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, et al. Whole-genome sequencing for national surveillance of Shiga toxin-producing Escherichia coli O157. Clin Infect Dis. 2015;61(3):305-12. https://doi.org/10.1093/cid/civ318 PMID: 25888672
7. Butcher H, Elson R, Chattaway MA, Featherstone CA, Willis C, Jorgensen F, et al. Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk. Epidemiol Infect. 2016;144(13):2812-23. https://doi.org/10.1017/S0950268816000509 PMID: 27338677
8. Mikhail AFW, Jenkins C, Dallman TJ, Inns T, Martín AIC, Fox A, et al. An outbreak of Shiga toxin-producing Escherichia coli O157:H7 associated with contaminated salad leaves: epidemiological, genomic and food trace back investigations. Epidemiol Infect. 2018;146(2):187-96. https://doi.org/10.1017/S0950268817002874 PMID: 29248018
9. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95. https://doi.org/10.1093/bioinformatics/btp698 PMID: 20080505
10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199
11. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623
12. Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, Shaaban S, et al. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. Bioinformatics. 2018;34(17):3028-9. https://doi.org/10.1093/bioinformatics/bty212 PMID: 29659710
13. Ashton PM, Perry N, Ellis R, Petrovska L, Wain J, Grant KA, et al. Insight into Shiga toxin genes encoded by Escherichia coli O157 from whole genome sequencing. PeerJ. 2015;3:e739. https://doi.org/10.7717/peerj.739 PMID: 25737808
14. Sadiq SM, Hazen TH, Rasko DA, Eppinger M. EHEC Genomics: Past, Present, and Future. Microbiol Spectr. 2014;2(4):0020-2013.
15. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol. 2014;52(5):1501-10. https://doi.org/10.1128/JCM.03617-13 PMID: 24574290
16. Holmes A, Allison L, Ward M, Dallman TJ, Clark R, Fawkes A, et al. Utility of Whole-Genome Sequencing of Escherichia coli O157 for Outbreak Detection and Epidemiological Surveillance. J Clin Microbiol. 2015;53(11):3565-73. https://doi.org/10.1128/JCM.01066-15 PMID: 26354815
17. Ferdous M, Friedrich AW, Grundmann H, de Boer RF, Croughs PD, Islam MA, et al. Molecular characterization and phylogeny of Shiga toxin-producing Escherichia coli isolates obtained from two Dutch regions using whole genome sequencing. Clin Microbiol Infect. 2016;22(7):642.e1-9. https://doi.org/10.1016/j.cmi.2016.03.028 PMID: 27058887
18. Parsons BD, Zelyas N, Berenger BM, Chui L. Detection, Characterization, and Typing of Shiga Toxin-Producing Escherichia coli. Front Microbiol. 2016;7:478. https://doi.org/10.3389/fmicb.2016.00478 PMID: 27148176
19. Rusconi B, Sanjar F, Koenig SS, Mammel MK, Tarr PI, Eppinger M. Whole Genome Sequencing for Genomics-Guided Investigations of Escherichia coli O157:H7 Outbreaks. Front Microbiol. 2016;7:985. https://doi.org/10.3389/fmicb.2016.00985 PMID: 27446025
20. Gilchrist CA, Turner SD, Riley MF, Petri WA Jr, Hewlett EL. Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev. 2015;28(3):541-63. https://doi.org/10.1128/CMR.00075-13 PMID: 25876885
21. Byrne L, Jenkins C, Launders N, Elson R, Adak GK. The epidemiology, microbiology and clinical impact of Shiga toxin-producing Escherichia coli in England, 2009-2012. Epidemiol Infect. 2015;143(16):3475-87. https://doi.org/10.1017/S0950268815000746 PMID: 25920912
22. Byrne L, Adams N, Glen K, Dallman TJ, Kar-Purkayastha I, Beasley G, et al. Epidemiological and Microbiological Investigation of an Outbreak of Severe Disease from Shiga Toxin-Producing Escherichia coli O157 Infection Associated with Consumption of a Slaw Garnish. J Food Prot. 2016;79(7):1161-8. https://doi.org/10.4315/0362-028X.JFP-15-580 PMID: 27357035
23. Buchholz U, Bernard H, Werber D, Böhmer MM, Remschmidt C, Wilking H, et al. German outbreak of Escherichia coli O104:H4 associated with sprouts. N Engl J Med. 2011;365(19):1763-70. https://doi.org/10.1056/NEJMoa1106482 PMID: 22029753
24. Wilson D, Dolan G, Aird H, Sorrell S, Dallman TJ, Jenkins C, et al. Farm-to-fork investigation of an outbreak of Shiga toxin-producing Escherichia coli O157. Microb Genom. 2018;4(3):e000160. PMID: 29488865
25. Rowell S, King C, Jenkins C, Dallman TJ, Decraene V, Lamden K, et al. An outbreak of Shiga toxin-producing Escherichia coli serogroup O157 linked to a lamb-feeding event. Epidemiol Infect. 2016;144(12):2494-500. https://doi.org/10.1017/S0950268816001229 PMID: 27297133
26. Strachan NJ, Rotariu O, Lopes B, MacRae M, Fairley S, Laing C, et al. Whole Genome Sequencing demonstrates that Geographic Variation of Escherichia coli O157 Genotypes Dominates Host Association. Sci Rep. 2015;5(1):14145. https://doi.org/10.1038/srep14145 PMID: 26442781
27. Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, Perry N, et al. Public health value of next-generation DNA sequencing of enterohemorrhagic Escherichia coli isolates from an outbreak. J Clin Microbiol. 2013;51(1):232-7. https://doi.org/10.1128/JCM.01696-12 PMID: 23135946
28. Jenkins C, Dallman TJ, Launders N, Willis C, Byrne L, Jorgensen F, et al. Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing Escherichia coli O157 Associated with Consumption of Watercress. Appl Environ Microbiol. 2015;81(12):3946-52. https://doi.org/10.1128/AEM.04188-14 PMID: 25841005
29. Launders N, Locking ME, Hanson M, Willshaw G, Charlett A, Salmon R, et al. A large Great Britain-wide outbreak of STEC O157 phage type 8 linked to handling of raw leeks and potatoes. Epidemiol Infect. 2016;144(1):171-81. https://doi.org/10.1017/S0950268815001016 PMID: 26041509
30. Sinclair C, Jenkins C, Warburton F, Adak GK, Harris JP. Investigation of a national outbreak of STEC Escherichia coli O157 using online consumer panel control methods: Great Britain, October 2014. Epidemiol Infect. 2017;145(5):864-71. https://doi.org/10.1017/S0950268816003009 PMID: 27964764
31. Gobin M, Hawker J, Cleary P, Inns T, Gardiner D, Mikhail A, et al. National outbreak of Shiga toxin-producing Escherichia coli O157:H7 linked to mixed salad leaves, United Kingdom, 2016. Euro Surveill. 2018;23(18):17-00197. https://doi.org/10.2807/1560-7917.ES.2018.23.18.17-00197 PMID: 29741151
32. Cowley LA, Dallman TJ, Fitzgerald S, Irvine N, Rooney PJ, McAteer SP, et al. Short-term evolution of Shiga toxin-producing Escherichia coli O157:H7 between two food-borne outbreaks. Microb Genom. 2016;2(9):e000084. PMID: 28348875
33. Franz E, Delaquis P, Morabito S, Beutin L, Gobius K, Rasko DA, et al. Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing Escherichia coli (STEC) in global food production systems. Int J Food Microbiol. 2014;187:57-72. https://doi.org/10.1016/j.ijfoodmicro.2014.07.002 PMID: 25051454
34. Allard MW, Bell R, Ferreira CM, Gonzalez-Escalona N, Hoffmann M, Muruvanda T, et al. Genomics of foodborne pathogens for microbial food safety. Curr Opin Biotechnol. 2018;49:224-9. https://doi.org/10.1016/j.copbio.2017.11.002 PMID: 29169072

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2019.

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

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

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

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

Folkhälsomyndighetens nyhetsbrev
Folkhälsomyndigheten, Stockholm
Weekly, online. In Swedish.
http://www.folkhalsomyndigheten.se/

United Kingdom

England and Wales

Health Protection Report
Public Health England, London
Weekly, online only. In English.
https://www.gov.uk/government/collections/health-protection-report-latest-infection-reports

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/

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/

European Centre for Disease Prevention and Control

European Centre for Disease Prevention and Control (ECDC)
The European Centre for Disease Prevention and Control (ECDC) was established in 2005. It is an EU agency with aim to strengthen Europe's defences against infectious diseases. It is seated in Stockholm, Sweden.
http://www.ecdc.europa.eu

All material in *Eurosurveillance* is in the public domain and may be used and reprinted without special permission. However, the source should be cited properly and we suggest adding a link to the exact page on the *Eurosurveillance* website.

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.

The opinions expressed by authors contributing to *Eurosurveillance* do not necessarily reflect the opinions of the European Centre for Disease Prevention and Control (ECDC) or the Editorial team or the institutions with which the authors are affiliated. Neither the ECDC nor any person acting on behalf of the ECDC is responsible for the use which might be made of the information in this journal.

TQ-AD-19-001-EN-C

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

The *Eurosurveillance* print edition is a compilation of 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 email, 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 : #eurosurveillanc

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

Paper TQ-AD-19-001-EN-C ISSN 1025-496X
PDF TQ-AD-19-001-EN-N ISSN 1560-7917

Graphic design © ECDC, Stockholm

Publications Office
of the European Union