EDITORIAL

How do advanced diagnostics support public health policy development?

Jacob Moran-Gilad

RAPID COMMUNICATIONS

Interim estimates of 2018/19 vaccine effectiveness against influenza A(H1N1)pdm09, Canada, January 2019

Danuta M Skowronska, Slobban Leir, Suzana Sabaiduc, Michelle Murti, James A Dickinson, Romy Olisha, Jonathan B Gubbay, Matthew A Croxen, Hugues Charest, Tracy Chan, Nathalie Bastien, Yan Li, Mel Krajden and Gaston De Serres

Genomic sequence of yellow fever virus from a Dutch traveller returning from the Gambia-Senegal region, the Netherlands, November 2018

My VT Phan, Sarwa Darwish Murad, Annemiek A van der Eijk, Herold J. Metselaar, Hermien Hartog, Femme Harinck, Corine H GeurtsvanKessel, Richard Molenkamp, Matthew Cotten and Marion PG Koopmans

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, Laura Pérez-Lago, Miguel Martínez-Lirola, Álvaro Chiner-Oms, Marta Herranz, Imane Chaoui, Ifaki Comas, My Driss El Messaoudi, José Antonio Garrido Cárdenas, Sheila Santantón, Emilio Bouza and Darío García-de-Yedema

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

Cecilia Fazio, Arianna Neri, Paola Vacca, Andrea Ciammaruconi, Milena Arghittu, Anna Maria Barbui, Caterina Vocale, Paola Bernaschi, Patrizia Isola, Irene Alessandra Galanti, Antonella Mencacci, Rosella De Nittis, Maria Chironna, Anna Giammanco, Elisabetta Pagani, Alessandro Bisbano and Paola Stefanelli

RESEARCH

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

Michelle S Toleman, Sandra Reuter, Dorota Jamrozy, Hayley J Wilson, Beth Blane, Ewan M Harrison, Francesc Coll, Russell J Hope, Angela Kearns, Julian Parkhill, Sharon J Peacock and M Estée Török

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

Lotta Silva, Umaer Naseer, Kristian Alfsnes, Nils Olav Hermansen, Heidi Lange and Lin T Brandal

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, Sylvia Gautsch, Ruth Spieler-Denz, Olivier Dubuis, Valeria Gaia, Jacob Moran-Gilad, Vladimira Hlinic, Helena MB Seth-Smith, Christian H. Nickel, Sarah Tschudin-Sutter, Stefano Bassetti, Monika Haenggi, Peter Brodmann, Simon Fuchs and Adrian Egli

REVIEW

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

Bélen Rodríguez-Sánchez, Emilia Cercenado, Alix T. Coste and Gilbert Greub

PERSPECTIVE


Claire Jenkins, Timothy J Hallman and Kathie A Grant
How do advanced diagnostics support public health policy development?

Jacob Moran-Gilad
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)

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. 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.

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. 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. 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...
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 Tolemen 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 pilgrimages [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 implication, 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 45%. 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


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
Interim estimates of 2018/19 vaccine effectiveness against influenza A(H1N1)pdm09, Canada, January 2019

Danuta M Skowronski¹,², Siobhan Leir¹, Suzana Sabaiduc¹, Michelle Murti¹,², James A Dickinson¹, Romy Olsha², Jonathan B Gubbay¹,², Matthew A Croxen³,⁴, Hugues Charest⁵, Tracy Chan¹, Nathalie Bastien³, Yan Li, Mel Krajden¹,², Gaston De Serres⁸,¹,¹¹

¹. British Columbia Centre for Disease Control, Vancouver, Canada
². University of British Columbia, Vancouver, Canada
³. Public Health Ontario, Toronto, Canada
⁴. University of Toronto, Toronto, Canada
⁵. University of Calgary, Calgary, Canada
⁶. Provincial Laboratory for Public Health, Edmonton, Canada
⁷. University of Alberta, Edmonton, Canada
⁸. Institut National de Santé Publique du Québec, Quebec City, Canada
⁹. National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada
10. Laval University, Quebec City, Canada
11. Centre Hospitalier Universitaire de Québec, Quebec City, Canada

Correspondence: Danuta M Skowronski (danuta.skowronski@bccdc.ca)

Using a test-negative design, the Canadian Sentinel Practitioner Surveillance Network assessed interim 2018/19 vaccine effectiveness (VE) against predominant influenza A(H1N1)pdm09 viruses. Adjusted VE was 72% (95% confidence interval: 60 to 81) against medically attended, laboratory-confirmed influenza A(H1N1)pdm09 illness. This substantial vaccine protection was observed in all age groups, notably young children who appeared to be disproportionately affected. Sequence analysis identified heterogeneity in emerging clade 6B.1 viruses but no dominant drift variant.

The 2018/19 influenza season in Canada for the period spanning November through January has been characterised by dominant influenza A(H1N1)pdm09 activity, with lesser influenza A(H3N2) and little influenza B contribution [1]. This profile is in contrast to the 2017/18 season which was comprised of dominant influenza A(H3N2) and early influenza B(Yamagata) co-circulation [2]. The last influenza A(H1N1)pdm09-dominant epidemics in Canada were in 2013/14 and 2015/16 [3,4].

The 2018/19 influenza vaccine for the northern hemisphere contains an A/Michigan/45/2015 (H1N1)pdm09-like antigen (belonging to clade 6B.1). The same component was included in the 2017/18 northern and the 2018 southern hemisphere vaccines [5]. Preliminary estimates of vaccine effectiveness (VE) from Australia’s 2018 season showed substantial VE of 78% (95% confidence interval (CI): 51 to 91) against influenza A(H1N1)pdm09 viruses [6]. Here we present interim 2018/19 VE estimates against influenza A(H1N1)pdm09 viruses from the Canadian Sentinel Practitioner Surveillance Network (SPSN), including detailed genetic characterisation of contributing viruses.

Vaccine effectiveness evaluation
VE was estimated using a test-negative design, as previously described [2-4]. Nasal/nasopharyngeal specimens and epidemiological data were collected from patients presenting to community-based sentinel practitioners in Alberta, British Columbia, Ontario, and Quebec. Influenza-like illness (ILI) was defined as acute onset of self-reported fever and cough and at least one other symptom including sore throat, myalgia, arthralgia or prostration. Fever was not a requirement for elderly adults 65 years and older. Analyses were restricted to patients at least 1-year-old presenting within 7 days of ILI onset. Vaccination status was based on self-report of 2018/19 vaccine receipt at least 2 weeks before symptom onset; patients vaccinated less than 2 weeks before onset or with unknown vaccination status or timing were excluded. Compliance with two-dose recommendations in young children was not assessed. All influenza vaccines manufactured for Canada (including SPSN provinces) for 2018/19 were
egg-based and the vast majority (> 95%) were inactivated vaccines (i.e. the live attenuated influenza vaccine constituted < 5% of doses distributed by the publicly-funded immunisation campaign). A high dose inactivated formulation was available for elderly adults in the SPSN province of Ontario. Institutional review boards in each province provided ethical approval.

Specimens collected from week 45 (starting 4 November 2018) to week 2 (ending 12 January 2019) were tested for influenza type and subtype by real-time RT-PCR assays at provincial public health reference laboratories. Sanger sequencing of the haemagglutinin (HA) gene (HA1 and HA2) of A(H1N1)pdm09 viruses contributing to VE analyses was conducted for all SPSN provinces at the British Columbia Centre for Disease Control Public Health Laboratory. Virus sequencing was undertaken on original patient specimens, including as many as possible in the order received. Viral sequence data are being deposited for reference into the Global Initiative on Sharing All Influenza Data (GISAID) platform (www.gisaid.org)

---

**Figure 1**

Influenza detections among eligible patients presenting with influenza-like illness, by week of specimen collection, Canadian Sentinel Practitioner Surveillance Network, 4 November 2018–12 January 2019 (n = 1,518)

---

*One specimen was co-infected with influenza A(H1N1)pdm09 and influenza B and another specimen was co-infected with influenza A(H1N1) pdm09 and A(H3N2). Viruses involved in co-infections are plotted separately by type/subtype such that specimens involved in co-infections appear twice.*

*Missing specimen collection dates were imputed as the date the specimen was received and processed at the provincial laboratory minus 2 days, the average time between specimen collection date and laboratory received date among specimens with complete information for both values. Data for week 2 may be incomplete given delays in specimen processing and laboratory testing.
### Table 1
Participant profile, by influenza A(H1N1)pdm09 case and vaccination status, Canadian Sentinel Practitioner Surveillance Network, 4 November 2018–12 January 2019 \( (n=1,442) \)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall Distribution by case status (column %)</th>
<th>Vaccination coverage (row %)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A(H1N1)pdm09 cases</td>
<td>Negative controls</td>
</tr>
<tr>
<td>n (row %)</td>
<td>n %</td>
<td>n %</td>
</tr>
<tr>
<td>Overall</td>
<td>n (1,442) 100</td>
<td>585</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–8</td>
<td>282</td>
<td>20</td>
</tr>
<tr>
<td>9–19</td>
<td>134</td>
<td>9</td>
</tr>
<tr>
<td>20–49</td>
<td>632</td>
<td>44</td>
</tr>
<tr>
<td>50–64</td>
<td>256</td>
<td>18</td>
</tr>
<tr>
<td>≥ 65</td>
<td>138</td>
<td>10</td>
</tr>
<tr>
<td>Median age (range)</td>
<td>35 (1–97)</td>
<td>31 (1–82)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>869</td>
<td>60</td>
</tr>
<tr>
<td>Male</td>
<td>564</td>
<td>39</td>
</tr>
<tr>
<td>Unknown</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Comorbidity(d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,099</td>
<td>76</td>
</tr>
<tr>
<td>Yes</td>
<td>264</td>
<td>18</td>
</tr>
<tr>
<td>Unknown</td>
<td>79</td>
<td>5</td>
</tr>
<tr>
<td>Province</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alberta</td>
<td>432</td>
<td>30</td>
</tr>
<tr>
<td>British Columbia</td>
<td>267</td>
<td>19</td>
</tr>
<tr>
<td>Ontario</td>
<td>546</td>
<td>38</td>
</tr>
<tr>
<td>Quebec</td>
<td>197</td>
<td>14</td>
</tr>
<tr>
<td>Specimen collection interval from ILI onset (days)(e)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 4</td>
<td>1,066</td>
<td>74</td>
</tr>
<tr>
<td>5–7</td>
<td>376</td>
<td>26</td>
</tr>
<tr>
<td>Median interval (range)</td>
<td>3 (0–7)</td>
<td>3 (0–7)</td>
</tr>
<tr>
<td>Month of specimen collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>November</td>
<td>409</td>
<td>28</td>
</tr>
<tr>
<td>December</td>
<td>736</td>
<td>51</td>
</tr>
<tr>
<td>January</td>
<td>297</td>
<td>21</td>
</tr>
<tr>
<td>Vaccination status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccination without regard to timing(f)</td>
<td>334/1,497</td>
<td>22</td>
</tr>
<tr>
<td>≥ 2 weeks before ILI onset</td>
<td>279</td>
<td>19</td>
</tr>
</tbody>
</table>

ILI: influenza-like illness; NA: not applicable.

Unless otherwise specified, values displayed in the columns represent the number of specimens per category and percentages are relative to the total. Where the denominator for the percentages differs from the total, fractions supporting the calculation of percentages are shown.

\(p\) values for comparison between cases and controls were derived by chi-squared test, Fisher's exact test or Wilcoxon rank-sum test.

Vaccination status based on patients’ self-report; defined as receipt of 2018/19 seasonal influenza vaccine at least 2 weeks before symptom onset. Patients vaccinated less than 2 weeks before onset of symptoms or with unknown vaccination status or timing were excluded.

\(p\) values for comparison of the proportion vaccinated were derived by chi-squared test, Fisher’s exact test or Wilcoxon rank-sum test.

Includes chronic comorbidities that place individuals at higher risk of serious complications from influenza as defined by Canada’s National Advisory Committee on Immunization, including: heart, pulmonary (including asthma), renal, metabolic (such as diabetes), blood, cancer or immunocompromising conditions, conditions that compromise management of respiratory secretions and increase risk of aspiration, or morbid obesity (body mass index ≥ 40).

\(p\) values for comparison between cases and controls were derived by chi-squared test, Fisher’s exact test or Wilcoxon rank-sum test.

Missing specimen collection dates were imputed as the date the specimen was received and processed at the laboratory minus 2 days, the average time between specimen collection date and laboratory received date among specimens with complete information for both values.

Participants who received seasonal 2018/19 influenza vaccine less than 2 weeks before ILI onset or for whom vaccination timing was unknown were excluded from the primary analysis. They are included here for assessing vaccination regardless of timing for comparison to other estimates of vaccination coverage.
### Table 2
Interim vaccine effectiveness estimates against any influenza infection, influenza A, and influenza A(H1N1)pdm09, Canadian Sentinel Practitioner Surveillance Network, 4 November 2018–12 January 2019 (n = 1,518)

<table>
<thead>
<tr>
<th>Model</th>
<th>Any influenza</th>
<th>Influenza A</th>
<th>Influenza A(H1N1)pdm09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n vac/N</td>
<td>%</td>
<td>n vac/N</td>
</tr>
<tr>
<td><strong>Primary analysis – all participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,518</td>
<td>1,514</td>
<td>1,442</td>
</tr>
<tr>
<td>Cases</td>
<td>59/661</td>
<td>9</td>
<td>58/657</td>
</tr>
<tr>
<td>Controls</td>
<td>234/857</td>
<td>27</td>
<td>234/857</td>
</tr>
<tr>
<td>Vaccine effectiveness</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>74</td>
<td>65 to 81</td>
<td>74</td>
</tr>
<tr>
<td><strong>Univariable adjustment for</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Age group (1–8, 9–19, 20–64, ≥ 65 years)</td>
<td>69</td>
<td>57 to 77</td>
<td>69</td>
</tr>
<tr>
<td>- Province (AB, BC, ON, QC)</td>
<td>73</td>
<td>63 to 80</td>
<td>73</td>
</tr>
<tr>
<td>- Interval from ILI onset to specimen collection (≤ 4, 5–7 days)</td>
<td>73</td>
<td>63 to 80</td>
<td>73</td>
</tr>
<tr>
<td>- Calendar timea</td>
<td>75</td>
<td>66 to 81</td>
<td>75</td>
</tr>
<tr>
<td>Full covariate adjustmentb</td>
<td>68</td>
<td>55 to 77</td>
<td>68</td>
</tr>
<tr>
<td><strong>Age-restricted analyses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants 1–8 years-old</td>
<td>n vac/N</td>
<td>%</td>
<td>n vac/N</td>
</tr>
<tr>
<td>Total</td>
<td>289</td>
<td>289</td>
<td>282</td>
</tr>
<tr>
<td>Cases</td>
<td>4/170</td>
<td>2</td>
<td>4/170</td>
</tr>
<tr>
<td>Controls</td>
<td>21/119</td>
<td>18</td>
<td>21/119</td>
</tr>
<tr>
<td>Vaccine effectiveness</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>89</td>
<td>66 to 96</td>
<td>89</td>
</tr>
<tr>
<td>Full covariate adjustmentc</td>
<td>88</td>
<td>60 to 96</td>
<td>88</td>
</tr>
<tr>
<td>Participants 9–19 years-old</td>
<td>n vac/N</td>
<td>%</td>
<td>n vac/N</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>138</td>
<td>134</td>
</tr>
<tr>
<td>Cases</td>
<td>2/58</td>
<td>3</td>
<td>2/58</td>
</tr>
<tr>
<td>Controls</td>
<td>9/80</td>
<td>11</td>
<td>9/80</td>
</tr>
<tr>
<td>Vaccine effectiveness</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>72</td>
<td>−36 to 94</td>
<td>72</td>
</tr>
<tr>
<td>Full covariate adjustmentc</td>
<td>71</td>
<td>−56 to 95</td>
<td>71</td>
</tr>
<tr>
<td>Participants 20–64 years-old</td>
<td>n vac/N</td>
<td>%</td>
<td>n vac/N</td>
</tr>
<tr>
<td>Total</td>
<td>946</td>
<td>943</td>
<td>888</td>
</tr>
<tr>
<td>Cases</td>
<td>41/401</td>
<td>10</td>
<td>40/398</td>
</tr>
<tr>
<td>Controls</td>
<td>133/545</td>
<td>24</td>
<td>133/545</td>
</tr>
<tr>
<td>Vaccine effectiveness</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>65</td>
<td>49 to 76</td>
<td>65</td>
</tr>
<tr>
<td>Full covariate adjustmentc</td>
<td>63</td>
<td>46 to 75</td>
<td>64</td>
</tr>
<tr>
<td>Participants ≥ 65 years-old</td>
<td>n vac/N</td>
<td>%</td>
<td>n vac/N</td>
</tr>
<tr>
<td>Total</td>
<td>145</td>
<td>144</td>
<td>138</td>
</tr>
<tr>
<td>Cases</td>
<td>12/32</td>
<td>38</td>
<td>12/32</td>
</tr>
<tr>
<td>Controls</td>
<td>71/113</td>
<td>63</td>
<td>71/113</td>
</tr>
<tr>
<td>Vaccine effectiveness</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>65</td>
<td>20 to 84</td>
<td>63</td>
</tr>
<tr>
<td>Full covariate adjustmentc</td>
<td>64</td>
<td>8 to 86</td>
<td>63</td>
</tr>
</tbody>
</table>

AB: Alberta; BC: British Columbia; CI: confidence interval; ILI: influenza-like illness; n vac: number vaccinated; ON: Ontario; QC: Quebec.

a Calendar time is based on week of specimen collection, modelled using cubic spline function with three equally spaced knots.

b Age group (1–8, 9–19, 20–49, 50–64, ≥ 65 years), province, specimen collection interval and calendar time.

c Province, specimen collection interval and calendar time.

d Age group (20–49, 50–64 years), province, specimen collection interval and calendar time.
Emerging genetic variants were assigned to genetic subgroups based on defining amino acid substitutions [7,8]. Substitutions were assessed for their involvement of HA1 antigenic sites (especially immuno-dominant Sa and Sb, but also Ca1, Ca2 or Cb), or as otherwise relevant to diversifying selection [9-11]. Sequence analysis was in relation to the cell-passaged A/Michigan/45/2015 vaccine reference strain and the corresponding egg-adapted version (A/Michigan/45/2015 X-275) as well as an alternate egg-adapted strain (A/Singapore/GP1908/2015 IVR-180) also used by manufacturers.

Odds ratios (OR) comparing influenza test positivity between vaccinated and unvaccinated participants were calculated using a logistic regression model, adjusted for age group, province, time from ILI onset to specimen collection, and specimen collection date. VE was derived as (1 − OR) × 100%. VE was estimated against influenza A(H1N1)pdm09 in the primary

### Table 3

Virological profile of influenza A(H1N1)pdm09 specimens contributing to interim 2018/19 vaccine effectiveness evaluation, Canadian Sentinel Practitioner Surveillance Network, 5 November 2018–4 January 2019 (n = 240)

<table>
<thead>
<tr>
<th>Nextstrain subgroup</th>
<th>Genetic clade&lt;sup&gt;a&lt;/sup&gt; with subgroup substitutions&lt;sup&gt;b&lt;/sup&gt;</th>
<th>British Columbia&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Alberta&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Ontario&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Quebec&lt;sup&gt;f&lt;/sup&gt;</th>
<th>Overall&lt;sup&gt;g&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
</tr>
<tr>
<td>6b1.A</td>
<td>6B.1 + I286V + L372V</td>
<td>0 0</td>
<td>2 2</td>
<td>0 0</td>
<td>1 3</td>
<td>3 1</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + K302T + I404M + E506D + N523S</td>
<td>10 13</td>
<td>4 4</td>
<td>10 38</td>
<td>10 29</td>
<td>34 14</td>
</tr>
<tr>
<td>6b1.A1</td>
<td>6B.1 + S183P + R455G + P282A + I298V + H126Y</td>
<td>0 0</td>
<td>0 0</td>
<td>1 4</td>
<td>0 0</td>
<td>1 0</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + N451T + P137S (Ca2)</td>
<td>0 0</td>
<td>0 0</td>
<td>1 4</td>
<td>0 0</td>
<td>1 0</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + N451T + T185I (Sb)</td>
<td>26 35</td>
<td>72 69</td>
<td>0 0</td>
<td>0 0</td>
<td>98 41</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + N129D + N260D + T185I (Sb)</td>
<td>2 3</td>
<td>7 7</td>
<td>5 19</td>
<td>16 47</td>
<td>30 13</td>
</tr>
<tr>
<td></td>
<td>T185I (Sb) + V479I</td>
<td>2 3</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 1</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + N129D + N260D + R205K (Ca1) + K443R</td>
<td>4 5</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>4 2</td>
</tr>
<tr>
<td>6b1.A2</td>
<td>6B.1 + S183P + N129D + N260D + R205K (Ca1) + R74I (Cb)</td>
<td>2 3</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 1</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + E235D (Ca1) + N260D + V520A + T185I (Sb)</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>6 18</td>
<td>6 3</td>
</tr>
<tr>
<td></td>
<td>6B.1 + S183P + E235D (Ca1) + N260D + V520A + N156D (Ca2) +</td>
<td>0 0</td>
<td>0 0</td>
<td>4 15</td>
<td>1 3</td>
<td>5 2</td>
</tr>
<tr>
<td></td>
<td>K160M (Sa) + T162K + S478N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6b1.A3</td>
<td>6B.1 + S183P + T120A</td>
<td>9 12</td>
<td>0 0</td>
<td>1 4</td>
<td>0 0</td>
<td>10 4</td>
</tr>
<tr>
<td></td>
<td>6B.1 + T120A + P183S (reversion)</td>
<td>6 8</td>
<td>3 3</td>
<td>0 0</td>
<td>0 0</td>
<td>9 4</td>
</tr>
<tr>
<td>6b1.A4</td>
<td>6B.1 + S183P + L233I + R74G (Cb)</td>
<td>5 7</td>
<td>2 2</td>
<td>0 0</td>
<td>0 0</td>
<td>7 3</td>
</tr>
<tr>
<td>6b1.A5</td>
<td>6B.1 + L161I (Sa) + I404M + N455T + D501E</td>
<td>9 12</td>
<td>15 14</td>
<td>4 15</td>
<td>0 0</td>
<td>28 12</td>
</tr>
</tbody>
</table>

<sup>a</sup> Unless otherwise stated in the subgroup substitutions, all 6B.1 viruses also bear the substitutions S74R (Cb), S164T (Sa) and I295V relative to the cell-passaged vaccine strain A/Michigan/45/2015 plus M209K and R223K (receptor-binding site) relative to the egg-adapted version (X-275). Compared to the alternate egg-passaged vaccine strain (A/Singapore/GP1908/2015 IVR-180), viruses additionally bore T120A (except in those subgroups already bearing T120A), plus M209K and A225G (receptor-binding site), the latter instead of R223K.

<sup>b</sup> Antigenic site substitutions are shown in bold, and the antigenic site follows in brackets. H1 numbering is based on influenza A(H1N1)pdm09 with the signal peptide removed. Genetic variants displayed here have been aligned with nextstrain subgrouping [8], recognising differences in numbering approaches (HA1 and HA2).

<sup>c</sup> HA sequences available for 75 of 87 (86%) A(H1N1)pdm09 viruses contributing to analyses, with specimen collection dates spanning 9 November 2018–4 January 2019.

<sup>d</sup> HA sequences available for 105 of 219 (48%) A(H1N1)pdm09 viruses contributing to analyses with specimen collection dates spanning 5 November–14 December 2018.

<sup>e</sup> HA sequences available for 26 of 179 (15%) A(H1N1)pdm09 viruses contributing to analyses with specimen collection dates spanning 12 November–7 December 2018.

<sup>f</sup> HA sequences available for 34 of 100 (34%) A(H1N1)pdm09 viruses contributing to analyses with specimen collection dates spanning 13 November–28 December 2018.

<sup>g</sup> HA sequences available for 240 of 585 (41%) A(H1N1)pdm09 viruses contributing to analyses with specimen collection dates spanning 5 November 2018–4 January 2019.
**Figure 2**

Distribution of clade 6B.1 variants by province, Canadian Sentinel Practitioner Surveillance Network, 5 November 2018–4 January 2019 (n = 240)

![Distribution of clade 6B.1 variants by province](chart.png)

- **Province, number of viruses and span of collection dates**
  - **BC (n=75)**: 9 Nov 2018 – 4 Jan 2019
  - **AB (n=105)**: 5 Nov – 14 Dec 2018
  - **ON (n=26)**: 12 Nov – 7 Dec 2018
  - **QC (n=34)**: 13 Nov – 28 Dec 2018
  - **Overall (n=240)**: 5 Nov 2018 – 4 Jan 2019

**Proportion (%)**

- 100%
- 80%
- 60%
- 40%
- 20%
- 0%

**Proportion, number of viruses and span of collection dates**

- **6B.1 + I286V + I372V** --- nextstrain 6b1.A
- **6B.1 + S183P + K302T + I404M + E506D + N523S** --- nextstrain 6b1.A
- **6B.1 + S183P + R45G + P282A + I298V +/- H126Y** --- nextstrain 6b1.A1
- **6B.1 + S183P + N451T + P137S (Ca2) +/- V173I** --- nextstrain 6b1.A1
- **6B.1 + S183P + N451T + T185I (Sb) +/- P137S (Ca2)** --- nextstrain 6b1.A1
- **6B.1 + S183P + N129D + N260D + T185I (Sb) +/- V479I** --- nextstrain 6b1.A2
- **6B.1 + S183P + N129D + N260D + R205K (Ca1) +/- K443R or R74I (Cb)** --- nextstrain 6b1.A2
- **6B.1 + S183P + E235D (Ca1) + N260D + V520A +/- T185I (Sb)** --- nextstrain 6b1.A3
- **6B.1 + S183P + T120A** --- nextstrain 6b1.A3
- **6B.1 + S183P + L233I + R74G (Cb)** --- nextstrain 6b1.A4
- **6B.1 + L161I (Sa) + I404M + N455T + D501E** --- nextstrain 6b1.A5

**AB: Alberta; BC: British Columbia; ON: Ontario; QC: Quebec.**

Based on Sanger sequencing of the haemagglutinin (HA) gene. H1 numbering is based on influenza A(H1N1)pdm09 with the signal peptide removed. Genetic variants displayed here have been aligned with nextstrain subgrouping [8], recognising differences in numbering approaches (HA1 and HA2).

Unless otherwise stated in the subgroup substitutions, all 6B.1 viruses also bear the substitutions S74R (Cb), S164T (Sa) and I295V relative to the cell-passaged vaccine strain A/Michigan/45/2015 plus M209K and R223K (receptor-binding site) relative to the egg-adapted version (X-275). Compared to the alternate egg-passaged vaccine strain A/Singapore/GP1908/2015 IVR-180), viruses additionally bore T120A (except in those subgroups already bearing T120A), plus M209K and A225G (receptor-binding site), the latter instead of R223K.
Epidemiological findings
Among 1,518 eligible specimens, 661 (44%) tested positive for influenza, including 656 (99%) influenza A, 4 (1%) influenza B and 1 (0.2%) influenza A/B coinfection. Among the 626 (95%) subtyped influenza A viruses, 585 (93%) were A(H1N1)pdm09 and 41 (7%) were A(H3N2) (Figure 1). There were two co-infections involving influenza A(H1N1)pdm09: one with influenza B and another with influenza A(H3N2).

Participant profiles are displayed in Table 1 for the 585 influenza A(H1N1)pdm09 cases and 857 test-negative controls included in primary VE analysis. Most (62%) participants were adults 20–64-years-old. However, significantly more influenza A(H1N1)pdm09 cases (28%) than controls (14%) were children 1–8-years-old (p < 0.001); that age group also had the highest influenza A(H1N1)pdm09 test positivity (58%; 163/282). Conversely, significantly fewer influenza A(H1N1)pdm09 cases than controls were 65 years or older (4% vs 13%; p < 0.001), and this age group exhibited the lowest influenza A(H1N1)pdm09 test positivity (18%; 25/138). Overall, 27% of controls but just 8% of influenza A(H1N1)pdm09 cases were considered vaccinated (p < 0.001) (Table 1).

After adjustment for relevant covariates, VE against any influenza, foremost driven by A(H1N1)pdm09 viruses, was 68% (95% CI: 55 to 77); for influenza A(H1N1)pdm09 alone, it was 72% (95% CI: 60 to 81) (Table 2). Estimates for influenza A(H1N1)pdm09 were similar in sensitivity analyses: with additional adjustment for sex and comorbidity, VE was 74% (95% CI: 61 to 82), and with restriction to specimens collected from 2 December 2018 (week 49), VE was 70% (95% CI: 55 to 8). By age group, adjusted VE estimates against influenza A(H1N1)pdm09 were similar in sensitivity analyses: with additional adjustment for sex and comorbidity, VE was 74% (95% CI: 61 to 82), and with restriction to specimens collected from 2 December 2018 (week 49), VE was 70% (95% CI: 55 to 8). By age group, adjusted VE estimates against influenza A(H1N1)pdm09 were 91% (95% CI: 67 to 98) in 1–8-year-old children, 71% (95% CI: 60 to 95) in 9–19-year-old children, 68% (95% CI: 51 to 80) in 20–64-year-old adults and 65% (95% CI: 1 to 88) in adults 65 years and older.

Virological findings
Sequencing of the HA gene was available for 240 (41%) of 585 influenza A(H1N1)pdm09 viruses. Collection dates of sequenced viruses spanned from 5 November 2018 to 4 January 2019 (Table 3, Figure 2). All sequenced viruses belonged to genetic clade 6B.1, to which the A/Michigan/45/2015 vaccine reference virus also belongs. However, all sequenced viruses additionally bore substitutions S74R (Cb), S164T (Sa) and I295V compared with the cell-passaged A/Michigan/45/2015 vaccine strain (except for nine viruses which showed continued drift at position 74). All viruses also bore additional substitutions M209K and R223K (receptor-binding site) attributed to egg adaptation mutations in the A/Michigan/45/2015 X-275 vaccine strain.

Beyond these shared substitutions, we observed heterogeneity among sequenced influenza A(H1N1)pdm09 viruses overall and by province, with no single subgroup dominating. Across subgroups, an S183P (non-antigenic site) substitution was found in 200 (83%) of 240 viruses and T185I (Sb) was found in 136 (57%) of 240 viruses. In Alberta, where there was an earlier epidemic peak [1], most viruses (72/105; 69%) belonged to a 6B.1 subgroup bearing T185I (Sb) with S183P and N451T substitutions (both non-antigenic sites); the same variant was also identified in a substantial proportion of viruses in British Columbia (26/75; 35%). In Ontario and Quebec, where fewer viruses contributed to sequence analysis, a different mix of subgroups was identified (Table 3, Figure 2).

Discussion
In this interim analysis, the 2018/19 influenza vaccine is estimated to have reduced the risk of medically attended influenza A(H1N1)pdm09 illness in Canada by 72%. This 2018/19 mid-season VE estimate against dominant influenza A(H1N1)pdm09 viruses is substantially higher than last reported in the mid-season analysis from Canada for the 2017/18 A(H3N2)-dominant influenza epidemic, for which VE against A(H3N2) viruses was below 20% (with a paucity of A(H1N1)pdm09 cases detected) [2].

Our 2018/19 VE estimate of 72% (95% CI: 60 to 81) against influenza A(H1N1)pdm09 viruses is comparable to a preliminary report from Australia using the same vaccine component for their 2018 season (78%) [6]. Both estimates are higher than reported in prior meta-analysis for influenza A(H1N1)pdm09 viruses (61%; 95% CI: 57 to 65) [12]. The Canadian SPSN estimate for 2018/19 is similar to mid-season estimates from our network during the last two A(H1N1)pdm09-dominant epidemics in 2013/14 and 2015/16 [3,4]. Of note, the 2013/14 epidemic peaked in January 2014, with comparable VE estimates at mid- and end-of-season analysis (74%; 95% CI: 58 to 83 and 71%; 95% CI: 58 to 80, respectively) [10]. Conversely, the mid-season VE estimate for 2015/16 was substantially higher than the end-of-season estimate (64%; 95% CI: 44 to 77 vs 43%; 95% CI: 25 to 57), a finding that may in part be explained by waning of immunity and the unusually delayed epidemic peak in March 2016 (after which half the cases were accrued) [13]. Similar to 2013/14, the current season’s epidemic may have already peaked nationally in Canada; however, there is regional variation in the timing and intensity of activity [1]. Differences in VE estimates at end-of-season analysis cannot be ruled out.

Globally, influenza A(H1N1)pdm09 viruses are in genetic flux, with substantial heterogeneity in circulating clade 6B.1 viruses, but no dominant drift (immunological escape) variant yet declaring a fitness advantage (as in 2015/16) [8]. Consistent with virus characterisation in Europe [7], all sentinel A(H1N1)pdm09 viruses sequenced here belonged to clade 6B.1 and bore
additional S74R (Cb), S164T (Sa) and I295V mutations. Across various genetic subgroups, most viruses (83%) also bore S183P. Although the latter is not within an antigenic site, the introduction of proline (a large aromatic ring) in such close proximity to antigenic site Sb could have structural effects. A slim majority of sequenced viruses (57%) in several subgroups bore T185I substitution. Amino acid 185 in antigenic site Sb first mutated during the 2010/11 season and became established in the A(H1N1)pdm09 population as clades 6 and 7, with S183P becoming dominant in subsequent seasons. The extent to which T185I substitution may now instead prevail warrants monitoring, especially when present alongside S183P as prominently identified in western Canada. Of note, antigenic site Sa has also evolved considerably over the past seven seasons (with mutations localised around amino acids 160–163). While the 2013/14 A(H1N1)pdm09 epidemic was dominated by a K163Q mutant and 2015/16 by the clade 6B.1 S162N variant, this season L161I and K160M substitutions have arisen, albeit infrequently (12% and 2% of viruses, respectively). These substitutions may be particularly relevant since they are located adjacent to an important, experimentally determined B-cell epitope [11]. Further, the amino acid at position 74 in site Cb continues to drift, with 4% of SPSN viruses carrying either the R74I or R74G substitution.

That interim VE estimates were not markedly affected by this genetic heterogeneity is consistent with findings from national surveillance systems in Canada, the United States (US) and Europe reporting few influenza A(H1N1)pdm09 viruses (<5%) manifesting antigenic drift from the A/Michigan/45/2015 vaccine strain [1,7,14,15]. In Canada, low-level antigenic distinction has been restricted to viruses bearing a substitution at position 156, within the immuno-dominant Sa site and previously recognised as influential on antigenicity and receptor-binding properties [16-18]. Mutations at position 156, however, were identified in just 2% of SPSN sequences so far this season. In combination, these findings reinforce the World Health Organization’s decision to retain the A/Michigan/45/2015 strain for the forthcoming 2019 southern hemisphere vaccine [5]. However, global monitoring for further evolution in circulating variants remains important to inform potential vaccine reformulation for subsequent seasons.

As in prior SPSN analyses [2-4], most participants in the current analysis were adults aged 20–64 years (62%). However, 1–8-year-old children appeared to be disproportionately affected, accounting for 28% of A(H1N1)pdm09 cases overall in our outpatient setting (despite comprising ca 9% of the underlying population and 14% of controls across SPSN provinces [19]). This paediatric involvement is also consistent with national surveillance findings for 2018/19 [1] but may be more pronounced in the current analysis than in previous A(H1N1)pdm09-dominated seasons. For example, 160 (30%) of 540 unvaccinated influenza A(H1N1)pdm09 cases in 2018/19 were 1–8-year-old children (Table 1) whereas their corresponding contribution was significantly lower in 2015/16 (32/237; 14%; p<0.001) and 2013/14 (32/259; 12%; p<0.001), despite more comparable representation among unvaccinated controls across the same mid-season analyses in 2018/19 (98/623; 16%) compared with 2015/16 (77/454; 17%; p=0.66) and 2013/14 (37/332; 11%; p=0.04) [3,4]. Overall, the median age of unvaccinated influenza A(H1N1)pdm09 cases was lower this season (29 years) than in 2015/16 (36 years; p<0.001) or 2013/14 (35 years; p<0.001) [3,4]. These age-related differences may reflect a greater proportion of children younger than 9 years in the current epidemic who were not yet born during prior H1 epidemics – notably the 2009 pandemic of nine years ago – with fewer opportunities to acquire immunity compared to older age groups.

Limitations of the current analysis include its observational design for which residual bias and confounding cannot be ruled out. Sample size was limited in age-stratified analyses, requiring cautious interpretation. In addition, characteristics of the 2018/19 influenza season have varied across the northern hemisphere. While the season in Canada commenced earlier than in recent years [1], notable influenza activity in the US and most of Europe was not observed until mid-December [14,15]. Further, while influenza A(H3N2) has accounted for less than 10% of subtyped A viruses in this study and nationally in Canada [1], influenza A(H3N2) has predominated in south-eastern regions in the US, and co-circulation has been observed in Europe, with influenza A(H3N2) accounting for about one third of subtyped influenza A detections [14,15]. It remains to be seen how varying virological and participant profiles will impact VE estimates elsewhere across the northern hemisphere.

Conclusions

Interim estimates from Canada for the 2018/19 northern hemisphere indicate substantial VE of ca 70% against influenza A(H1N1)pdm09. Thus far, this epidemic has taken a greater toll on children younger than 9 years even when compared with previous A(H1N1)pdm09-dominant seasons. Given ongoing epidemic activity in some regions, vaccination should be advocated to minimise the A(H1N1)pdm09-associated disease burden. In the context of observed genetic diversity, monitoring for further evolution in circulating 6B.1 variants, and potential impact on vaccine protection, is warranted.

Acknowledgements

The authors gratefully acknowledge the contribution of sentinel sites whose regular submission of specimens and data provide the basis of our analyses. We wish to acknowledge the administrative, coordination and data entry support in each participating province including: Shinhye Kim for the British Columbia Centre for Disease Control, Dylan Kendrick, Yvonne Efegoma and Sarah McDonald for TARRANT in Alberta, Kathleen Parris for Public Health Ontario and Sophie Auger for the Institut national de santé publique du Québec. We also thank Yayuk Joffres of the British Columbia Centre for Disease Control for her contribution to the analyses. We
also wish to thank those who provided laboratory and technical support in each province at the British Columbia Centre for Disease Control Public Health Laboratory, the Alberta Provincial Laboratory for Public Health (ProLab), the Public Health Ontario Laboratory and the Laboratoire de santé publique du Québec (LSPQ). Finally, we acknowledge the GiSAID EpiFlu Database for serving as a repository for sharing viral genetic data and facilitating related analyses.

Funding was provided by the British Columbia Centre for Disease Control, Alberta Health and Wellness, Public Health Ontario, Ministère de la santé et des services sociaux du Québec and the Public Health Agency of Canada. The views expressed herein do not necessarily represent the view of the Public Health Agency of Canada.

Conflict of interest

GDS has received grants for investigator-initiated studies unrelated to influenza vaccine from Pfizer and provided paid expert testimony for the Ontario Nurses Association, the Quebec Ministry of Justice and GSK. JBG has received research grants from Pfizer Inc. to conduct microbiological surveillance of Streptococcus pneumoniae. MK has received research grants from Roche, Siemens, and Hologic for unrelated studies. Other authors have no conflicts of interest to declare.

Authors’ contributions

Principal investigators (epidemiological): DMS (National and British Columbia), JAD (Alberta), MM (Ontario) and GDS (Québec). Principal investigator (laboratory): MK (British Columbia), MC (Alberta), JBG (Ontario), HC (Québec) and NB and YL (National Microbiology Laboratory). Additional laboratory and epidemiological support: RO (Ontario). Genomic sequencing and analysis: SS and TC. Epidemiological data and laboratory and epidemiological support: RO (Ontario). Genomic genetic data and facilitating related analyses: SL and DMS. Preparation of first draft: SL and DMS. Draft revision and approval: all.

References


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
Genomic sequence of yellow fever virus from a Dutch traveller returning from the Gambia-Senegal region, the Netherlands, November 2018

My VT Phan¹, Sarwa Darwish Murad², Annemiek A van der Eijk¹, Herold J. Metselaar², Hermien Hartog³, Femme Harinck², Corine H GeurtsvanKessel¹, Richard Molenkamp¹, Matthew Cotten¹, Marion PG Koopmans¹

1. Department of Viroscience, Erasmus MC, Rotterdam, the Netherlands
2. Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam, the Netherlands
3. Division of Hepato-Pancreato-Biliary and Transplant Surgery, Department of Surgery, Erasmus MC, Rotterdam, the Netherlands

Correspondence: Marion P.G. Koopmans (m.koopmans@erasmusmc.nl)

Citation style for this article:

In November 2018, yellow fever was diagnosed in a Dutch traveller returning from a bicycle tour in the Gambia-Senegal region. A complete genome sequence of yellow fever virus (YFV) from the case was generated and clustered phylogenetically with YFV from the Gambia and Senegal, ruling out importation into the Netherlands from recent outbreaks in Brazil or Angola. We emphasise the need for increased public awareness of YFV vaccination before travelling to endemic countries.

We report the genomic sequence of yellow fever virus (YFV) genome directly from clinical samples from an unvaccinated Dutch traveller returning from the Gambia-Senegal region, where yellow fever (YF) is endemic. This report sends a reminder of the importance of vaccination for travellers to endemic areas and furthermore shares with the community a YFV genome sequence identified from an area with limited YFV sequence coverage.

Case description
The case was a healthy, unvaccinated adult in his mid-20s who had travelled to the Gambia and Senegal for a 17-day bicycle tour in November 2018 (14 days in the Gambia and 3 days in Senegal). The patient had reported insect bites while travelling. During the returning flight to the Netherlands on 17 November, the patient developed fever and chills and then quickly progressed to acute kidney injury and fulminant liver failure for which he was hospitalised 20 November [1]. On 21 November, the patient was referred to the Erasmus Medical Centre (Rotterdam, the Netherlands) for treatment. Based on the clinical presentation and the recent travel history, YF was suspected and confirmed by a real-time PCR diagnostic assay on samples collected on 19 November and confirmed again on samples collected on 21 November. The patient was discharged 3 weeks after admission and has fully recovered from the infection. Full details of the clinical course and the advanced treatment will be described elsewhere.

Sample processing and agnostic deep sequencing
An in-house standard PCR for YFV yielded a Ct-value of 14 for a plasma sample collected on 19 November. This sample was prepared for whole genome sequencing as follows. Total nucleic acid was extracted using Roche MagNa Pure high performance extraction kit (Roche, Mannheim, Germany), followed by reverse transcription using random hexamer primers that avoid rRNA binding. Second strand synthesis was performed as previously described [2], followed by standard Ion Torrent library preparation as per manufacturer’s instruction. Deep sequencing was performed on the S5-XL sequencer, generating ca 10 million short reads of median length 263 nt. Short and low quality reads (<75 nt, Phred score <25) were removed and the remaining reads were de novo assembled to larger contigs using SPAdes v.3.13.0 [3]. The YFV sequence contigs were identified using Usearch [4] against a set of viral family protein databases. A complete YFV genome (10771 nt) was obtained from the analysis.

Alignment and phylogenetic analysis
This YFV genome (GenBank accession number MK292067) and all available YFV genomes retrieved from GenBank (n=188) were aligned using MUSCLE [5], manually checked in AliView [6], and trimmed to the complete Open Reading Frame (ORF). The evolutionary model testing was implemented in IQ-TREE [7] using the Akaike Information Criterion (AIC).
Figure
Maximum-likelihood phylogenetic tree of the complete yellow fever virus genomes including sequence from Dutch traveller to the Gambia and Senegal in November 2018

The genotypes of each YFV clade (as defined in [11]) are shown and abbreviated. The phylogeny was mid-point rooted for clarity and only bootstrap values for major clades were shown. The scale bar is shown in units of number of nt substitutions per site (subs/site).
A maximum-likelihood phylogenetic tree was constructed using the sequence alignment in RAxML [8] under the GTR+Γ model of evolution, which was determined as the best-fitted model, bootstrapped with 100 pseudoreplicates. The resulting tree was visualised and edited in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and mid-point rooted for clarity.

Clustering with other YFV sequences
The reported YFV genome was found to belong to the West Africa genotype according to a genotyping tool (http://krisp.ukzn.ac.za/app/typingtool/yellowfevervirus/) and in a maximum-likelihood phylogenetic tree (Figure). The reported genome was most closely related to a Gambian YFV genome from 2001 [9] with 98.3% nt identity across the entire genome and 195 nt differences and to Senegalese YFV genomes identified in 2000 [10]. Earlier Senegalese YFV genomes from 1995, 1996, 2001 and 2005 belonged to related but distinct lineages within the West Africa genotype.

The viral sequence from the patient was clearly distinct from viral sequences from the recent large outbreaks in Brazil (SA1 lineage, Figure [11]) and Angola (EAf r lineage, Figure [12]), suggesting that the YFV infection was likely a sylvatic case derived from locally circulating viruses in the Gambia and Senegal and not a new introduction of the virus into this region. However, it should be noted that there is a paucity of publicly available YFV genome sequences from Africa.

Discussion and conclusion
YF is a severe, mosquito-borne flavivirus infection caused by YFV, that is estimated to result in 78,000 deaths annually in Africa alone [13,14]. YFV transmission continues in tropical regions of the world with larger recent outbreaks reported in Brazil [11] and Angola. A smaller number of cases are reported from additional countries as listed on World Health Organization (WHO) news on disease outbreak [15]. Given the identification of co-circulating YFV lineages in regions over several years [9,11,15-22] and the general lack of sampling in the animal reservoir, it is plausible that more diversity may be observed with more comprehensive sequencing of newly diagnosed cases. Such surveillance in this part of the world would provide further knowledge and understanding of YFV transmission and evolution, which would be valuable in supporting the YF epidemic elimination initiative.

Although an effective and safe vaccine has been available since 1939 [23], vaccine coverage is still insufficient and a limited vaccine supply coupled with human population increases has led to high numbers of unvaccinated people living in endemic regions [24]. There have been several reports of YF cases in unvaccinated travellers returning from endemic regions in the past years such as to Belgium from the Gambia [9,16,18], to China from Angola [17,19] and to the Netherlands from Suriname [20,21] or from Brazil [22]. Furthermore, returning travellers may serve as sentinels for local outbreaks of pathogenic viruses that may have not yet been documented or adequately reported.

The WHO has launched a programme to eliminate YF epidemics in regions at risk for cases from enzootic circulation or new introductions [25]. A key component of a successful elimination campaign is the ability to detect new cases and to understand the ecology of YF in regions at risk. Whole genome viral sequences can provide important data for tracking viruses within and between outbreaks [11,26-28]. Having a rapid whole-genome confirmation of a YF infection and placing the sequence in the context of the global YFV phylogenetics is crucial for ruling out alternate transmission possibilities such as importation and introduction of YFV into the Netherlands from the recent large YFV outbreaks in Brazil or Angola. This work also highlights the need to remain alert for unexpected infectious disease aetiologies in returning travellers and the need to consider vaccination before travelling to regions where YFV is endemic, even if the vaccination is not required by border control agencies or when there are no reports of human cases of YF in these regions.

Data availability
The YFV genomic sequence reported here is available on GenBank with the accession number MK292067.

Acknowledgements
We would like to thank the patient consent giving samples for research purposes. We appreciated all treating clinicians who helped take good care of the patient. We thank Ronald van Marion and Winand Dinjens (Department of Pathology, Erasmus MC, Rotterdam, the Netherlands) for their sequencing, assembled the genome, performed the phylogenetic analyses and prepared the first draft of the manuscript. All co-authors were involved in writing and revising the manuscript.

Conflict of interest
None declared.

Authors’ contributions
Sarwa Darwish Murad, Herold J. Metselaar, Hermien Hartog and Femme Harinck were responsible for patient care. Annemieke A. van der Elk, Corine H. Geurts van Kessel, Richard Molenkamp and Marion P.G. Koopmans were responsible for the yellow fever diagnostics and sample logistics. Marion P.G. Koopmans coordinated the entire effort and secured funding. My V.T. Phan and Matthew Cotten performed the sequencing, assembled the genome, performed the phylogenetic analyses and prepared the first draft of the manuscript. All co-authors were involved in writing and revising the manuscript.

References
1. ProMED-mail. PRO/AH/EDR> Yellow fever - Africa (22): Nigeria (ED), Netherlands ex Gambia/Senegal. Archive number
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 1, Álvaro Chiner-Oms 4, 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:

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,
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 assignment 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 1](chart.png)

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.
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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 included the maximum likelihood–inferred ancestral nt 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 [39]. 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 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 specified for isolates 113 and the other two alleles expected for non-113 strains) (Supplementary Table S2). 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 MgCl2, 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 specified 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

**Eastern Europe**

Cluster 951

Cluster 691

Cluster 74

Cluster 348

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.
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 shared by the two cases in the recent transmission group. 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 51: 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**

Morocco

Morocco

Morocco

Morocco

Morocco

**Cluster 005**

Morocco

Morocco

Morocco

Morocco

Romania

**Cluster 113**

Morocco

(2016)

(2014)

(2015)

(2011)

(2015)

(2016)

(2007)

Spain

Spain

Spain

Morocco

(2010)

(2003)

Morocco

(2015)

**Cluster 1192**

Morocco

Morocco

Morocco

Romania

Spain

---

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 homogygosis in the remaining cases. The years of diagnosis are indicated in brackets.
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 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.
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 autochthonous, 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 toanalyse 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
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 transmission 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 coexist 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 Escrivano (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; ELAC2015/018-0064), 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


License 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.

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 Fazio1, Arianna Neri1, Paola Vaccà1, Andrea Ciammaruconi2, Milena Arghittù1, Anna Maria Barbui4, Caterina Vocale1, Paola Bernaschi1, Patrizia Isola1, Irene Alessandra Galanti2, Antonella Mencacci3, Rosella De Nittis11, Maria Chironna3, Anna Giammanco10, Elisabetta Pagani13, Alessandro Bisbano14, Paola Stefanelli1
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)

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 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/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’).
[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 *Neisseria 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.
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.
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).

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

BAST: Bexsero antigen sequence typing; cgMLST: core genome multilocus sequence typing; ID: identification code.
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 MIC90 of 0.064 mg/L and 0.19 mg/L, respectively.

Moreover, 14 showed decreased susceptibility to cefotaxime, ceftriaxone, ciprofloxacin and rifampicin. All of them were susceptible to

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 cc22 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.
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.

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.

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 Dapraí (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
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 Arghitii, Anna Maria Barbui, Caterina Volcata, Paola Bernaschi, Patrizia Isola, Alessandra Irene Galanti, Antonella Mencacci, Rosella De Nittis, Maria Chironna, Anna Giammancio, Elisabetta Pagani, Alessandro Bisbano were involved in the invasive meningococcal disease 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


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
Prospective genomic surveillance of meticillin-resistant *Staphylococcus aureus* (MRSA) associated with bloodstream infection, England, 1 October 2012 to 30 September 2013

Michelle S Toleman¹,²,³, Sandra Reuter⁴, Dorota Jamrozy², Hayley J Wilson¹, Beth Blane¹, Ewan M Harrison¹,², Francesc Coll⁵, Russell J Hope⁶, Angela Kearns⁶, Julian Parkhill², Sharon J Peacock¹,²,³,⁵, M Estée Török¹,³,⁷

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
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:

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*
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), SCCmec-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-o) 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].
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 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 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

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\(^c\).

[Graph showing diversity of lineages (CC) within three isolate collections: carriage and clinical samples from the Cambridgeshire study of MRSA; the national PHE BSI collection and a national BSAC BSI collection.]
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 from every positive case (carrier 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 haematology 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.
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.
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 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 BSI**
- **Cambridge**
- **USA**
- **FPR3757 reference**

- **~100 SNPs**

---


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).
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).


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.
A further potential benefit of prospective sequencing of MRSA bloodstream isolates 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 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 CC55 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 G0000803) 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 HIW performed the laboratory work in Cambridge. RH and AK co-ordinated the collection and characterisation of isolates and provided the


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
Whole genome sequencing of *Salmonella* Chester reveals geographically distinct clusters, Norway, 2000 to 2016

Lotta Siira1,2, Umaer Naseer1, Kristian Alfsnes1, Nils Olav Hermansen1,3, Heidi Lange1, Lin T Brandal1

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)


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 *Salmonella enterica* 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 *Salmonella enterica* 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.
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 > 50×. Quality control of the raw reads was done through FastQC. The sequences were submitted to the European Nucleotide Archive (ENA) under the accession 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 1.2.04 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>
<thead>
<tr>
<th>WGS cluster*</th>
<th>ST</th>
<th>Travel history</th>
<th>Year</th>
<th>Resistance determinants identified through ResFinder</th>
<th>PlasmidFinder and pMLST result</th>
<th>Phenotypic antimicrobial susceptibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>ST411 (n = 14)</td>
<td>Greece (n = 8)</td>
<td>2016</td>
<td>QnrB19</td>
<td>IncI1, IncX3(pEC14), Col(pVC)</td>
<td>BL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>QnrB19</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>QnrB19</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>aph(6)-Id</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2006</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>QnrB19</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>ST5260 (n = 3)</td>
<td>Cyprus (n = 3)</td>
<td>2010</td>
<td>None</td>
<td>IncI2</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2014</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, floR, sul2, tet(A), dfrA14</td>
<td>Col(pVC), IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, floR, sul2, tet(A), dfrA14</td>
<td>Col(pVC), IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, floR, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2014</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, sul2, tet(A), dfrA14</td>
<td>Col(pVC), IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2014</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>aph(6)-Id, aph(3&quot;)-Ib sul2, tet(A), dfrA14</td>
<td>IncI1, ST21</td>
<td>S</td>
</tr>
<tr>
<td>Africa</td>
<td>ST1954 (n = 15)</td>
<td>Morocco (n = 14)</td>
<td>2013</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrS1, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>aph(6)-Id, aph(3&quot;)-Ib, QnrB19, sul2, tet(A), dfrA14</td>
<td>IncN_ST7</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>QnrB19</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>QnrB19</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>QnrB19</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>aph(6)-Id, aph(3&quot;)-Ib sul2, tet(A), dfrA14</td>
<td>IncI1, ST21</td>
<td>S</td>
</tr>
</tbody>
</table>

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.

* 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>
<thead>
<tr>
<th>WGS cluster</th>
<th>ST</th>
<th>Travel history</th>
<th>Year</th>
<th>Resistance determinants identified through ResFinder</th>
<th>PlasmidFinder and pMLST result</th>
<th>Phenotypic antimicrobial susceptibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL</td>
</tr>
<tr>
<td>Asia 1 (n = 11)</td>
<td>ST343 (n = 11)</td>
<td>Thailand (n = 6)</td>
<td>2008</td>
<td><em>bla</em>&lt;sub&gt;TEM&lt;/sub&gt;-1B, <em>QnrS1</em>, <em>gyrA</em>.S83F</td>
<td>IncN</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2008</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2009</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
<td><em>gyrA</em> p.S83F</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2013</td>
<td><em>aadA1</em>, <em>sul3</em></td>
<td>IncI&lt;sub&gt;1&lt;/sub&gt; - ST3</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None (n = 5)</td>
<td>2003</td>
<td><em>bla</em>&lt;sub&gt;TEM&lt;/sub&gt;-1B</td>
<td>IncN</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td>Asia 2 (n = 8)</td>
<td>ST2063 (n = 8)</td>
<td>Sri Lanka (n = 4)</td>
<td>2011</td>
<td><em>bla</em>&lt;sub&gt;TEM&lt;/sub&gt;-1B</td>
<td>IncL/M (<em>pOXA-48</em>)</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td><em>aph(6)-Id</em>, <em>blaTEM-1B</em>, <em>QnrS1</em>, <em>sul3</em>, tet(A), <em>dfrA14</em></td>
<td>IncX1</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2012</td>
<td>None</td>
<td>None</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td><em>QnrB19</em></td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thailand (n = 2)</td>
<td>2015</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2016</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not known (n = 1)</td>
<td>2014</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2015</td>
<td>None</td>
<td>Col(pVC)</td>
<td>S</td>
</tr>
</tbody>
</table>

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.

* 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.
inferred from 1,000 replicates was also produced based on SNP calling, and a bootstrap consensus tree model was used to infer phylogeny from the reference-maximum likelihood method based on the Tamura-Nei distance (R < 30 mm), NAL (R < 16 mm) and TET (R < 17 mm) 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].

Petersburg University, St. Petersburg, Russia) [20,21].

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

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].

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 (ST954); 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 resistance 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 blaTEM-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 sul2 and dfrA14 genes. Thirty 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 blaTEM-1B and qnrS1 genes. In the Asia 2 cluster, one isolate carried the IncX1 plasmid and the aph(6)-Id (strB), blaTEM-1B, qnrS1, sul2, 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, IncI/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 ST1411 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.
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.

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 intermediate 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 multiscource 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**

5. Centers for Disease Control and Prevention (CDC). Multistate outbreak of salmonella chester infections associated with


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
**Air-conditioner cooling towers as complex reservoirs and continuous source of *Legionella pneumophila* infection evidenced by a genomic analysis study in 2017, Switzerland**

Daniel Wüthrich1,2,3, Sylvia Gautsch4, Ruth Spieler-Denz5, Olivier Dubuis4, Valeria Gaia7, Jacob Moran-Gilad8,9, Vladimira Hinic1, Helena MB Seth-Smith4,5,6, Christian H. Nickel4,6, Sarah Tschudin-Sutter1,4, Stefano Bassetti12, Monika Haenggi4, Peter Brodmann4, Simon Fuchs5, Adrian Egli1,2

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. Viöllier, 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 supplies 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 (ELDSSNet) 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 bronchoalveolar lavages, were cultured for a maximum of 10 days at 36°C under 5% CO₂ 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 bromoresol 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 S1) 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
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 populations.
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 benefited 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, JM, HSS, CN, STS, SB, MH, PB, SF and AE wrote the manuscript.

References


License 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.
This article is copyright of the authors or their affiliated institutions, 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.
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’, ‘antimicrobial susceptibility testing’, ‘resistance mechanism’, ‘typing’, ‘identification’ or ‘diagnosis’. 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 causative 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 bacteraemia 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. abscessus 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 (Zn²⁺) 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 Zn²⁺ [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
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

---

**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.
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 powerful 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 (CPT14/00220) 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 organization, 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


License 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.

This article is copyright of the authors or their affiliated institutions, 2019.
Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing

Claire Jenkins¹, Timothy J Dallman¹, Kathie A Grant¹
¹. National Infection Service, Public Health England, United Kingdom
Correspondence: Claire Jenkins (claire.jenkins@phe.gov.uk)

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 the 1990s 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>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
<th>Benefits of the WGS approach and context</th>
</tr>
</thead>
</table>
| Butcher et al. 2016  | 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  | 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  | 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    | 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   | 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   | 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| 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    | 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   | 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.
Cases 1, 2, 3, 8 and 9, initially identified by epidemiological links (reporting the consumption of RDM from the same farm), are designated ‘W’. Cases 10–14 were identified by analysis of the WGS data (and subsequently found to have consumed RDM from the same farm), are designated ‘E’. Cases 4–7 initially identified by epidemiological links with 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 about their consumption of raw drinking milk; those cases who consumed RDM were reclassified as cases 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.

**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...
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 a 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 Ib and certain clades and clusters in sub-lineages Ic and I/II are associated with UK strains, whereas sub-lineages Ia, Ib and Ila 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

---

**Table 2**


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

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