The contribution of whole-genome sequencing to our understanding of the epidemiology and control of methicillin-resistant *Staphylococcus aureus*

Hilary Humphreys[a,b], David C. Coleman[c]

[a]Department of Clinical Microbiology, the Royal College of Surgeons in Ireland and the [b]Department of Microbiology, Beaumont Hospital, Dublin Ireland.

[c]Microbiology Research Unit, Dublin Dental University Hospital, Trinity College Dublin, University of Dublin, Dublin, Ireland

Address and contact details:

Department of Clinical Microbiology
RCSI Education and Research Centre
Beaumont Hospital
Dublin D09 YD 60, Ireland

'Phone +3531 8093710/3708
Fax +3531 8092871
Email hhumphreys@rcsi.ie

Running title: Whole genome sequencing and MRSA Control

Abstract

In recent years, approaches to tracking the spread of methicillin-resistant *Staphylococcus aureus* (MRSA) as part of outbreak management have used conventional DNA-based methods including pulsed field gel electrophoresis (PFGE) and *spa* typing. However, when a predominant clone is present, these methods may be insufficiently discriminatory. We conducted a literature search to highlight how whole genome sequencing (WGS) has revolutionised the investigation of outbreaks of MRSA, including intra-hospital spread and MRSA in the community, and to review its future potential. Whole genome sequencing provides enhanced isolate discrimination, as it permits the entire genomic DNA sequence of isolates to be rapidly determined and compared. Many software packages used for the analysis of WGS data are becoming increasingly available. To date WGS has been more sensitive in confirming outbreaks, often persisting for prolonged periods, previously undetected by conventional molecular typing. The evolving dynamic of spread from the community to hospitals, within and between hospitals, and from hospitals to the community, is only becoming clear with WGS studies, and is more complex and convoluted than widely appreciated. Also, WGS can exclude cross-transmission, when isolates are different. The challenges now are to make WGS technology more amenable for routine use and to develop an evidence-based consensus for sequence difference thresholds for isolates that they are deemed part of the same outbreak, including protracted outbreaks. Using such data in a timely way will provide increased sensitivity in detecting cross-transmission events earlier with the potential of preventing outbreaks to positively impact on infection prevention and control.

Keywords: Whole genome sequencing, MRSA, bioinformatics, MRSA clones, outbreaks, spread, acute hospitals, community, long stay residential units.

Background

Staphylococcus aureus is a common cause of healthcare-associated infection (HCAI). In a recent European study in children, *S. aureus* was responsible for 11% of infections and was second only to coagulase-negative staphylococci as a cause [1]. In a four-country prevalence survey of HCAI in the UK and Ireland, methicillin-resistant *S. aureus* (MRSA) was responsible for 15.8% of all HCAIs [2]. In addition to causing serious infection in acutely ill patients in hospital, MRSA can spread to and between residents in long stay residential units (LSRU) and day units, as increasingly the elderly population is managed outside acute hospitals.

Infections caused by MRSA have been prevalent in many countries for decades even if in recent years, the prevalence of some serious infections, such as bloodstream infections (BSI), have declined [3,4]. Data from the UK and Northern Ireland from 2010-2014 derived from mandatory reporting of *S. aureus* bloodstream infection has shown that the proportion of BSI due to *S. aureus* has fallen from 16.2% to 8.9% with a downward trend in all four countries [3]. Molecular typing is important to track spread, indicate the possible origins of outbreaks, and confirm the efficacy of outbreak control measures and track the emergence or evolution of new clones in the outbreak and endemic setting.

Conventional molecular typing of MRSA

Routes of MRSA transmission and spread in healthcare settings have traditionally been investigated using targeted DNA-based typing methods including pulse-field gel electrophoresis (PFGE), *spa* typing, conventional multilocus sequence typing (MLST), and more recently, the application of DNA microarry technology [5,6,7]. In many cases these approaches provide valuable insights into the relatedness of clinical isolates when combined with relevant

74 epidemiological data and the expertise and insights of infection prevention and control staff.
75 However, conventional molecular typing approaches frequently struggle to discriminate between
76 isolates in the healthcare setting where a particular or limited number of MRSA clonal lineages
77 predominate. In many countries MRSA responsible for HCAIs frequently belong to a relatively
78 small number of clones; for example, in Ireland and the UK multilocus sequence type (ST) 22
79 has predominated among MRSA responsible for bloodstream infections and carriage for almost
80 two decades [7-9]. Isolates belonging to this lineage are particularly recalcitrant to discrimination
81 by conventional molecular typing approaches, making detection of all or even major patterns of
82 spread problematic [7, 10,11]. Consequently, there is a need for more discriminatory methods to
83 distinguish clonal isolates to track spread.
84
85 **Brief overview of whole genome sequencing of nosocomial pathogens**
86 *Next generation sequencing*
87 Table 1 provides a list and explanations of commonly used terms relating to WGS and its
88 analysis. Over the last decade the development of next generation sequencing (NGS) to
89 determine the entire genetic sequence of microbial pathogens, especially bacterial species, has
90 revolutionised molecular epidemiology. In contrast to Sanger sequencing, NGS permits the high-
91 throughput and rapid determination of whole genome sequences of pathogens at an affordable
92 cost. A variety of NGS approaches have been developed that utilise innovative sequencing
93 chemistry methods in tandem with small footprint automated bench top sequencers [12-15].
94 These approaches involve fragmentation of genomic DNA into short segments of a few hundred
95 nucleotide bases in length, tagging of the fragments with adapters to generate genome fragment
96 libraries, immobilisation of the libraries on a solid interface, followed by PCR amplification of

the DNA fragments. The DNA sequence of each fragment is determined during complementary strand synthesis using a variety of approaches depending on the NGS platform being used. This allows millions to billions of DNA fragments to be sequenced in parallel.

Two of the most widely used NGS platforms, also known as second generation sequencers or short read sequencers, include the Illumina sequencing systems (Illumina, Eindhoven, The Netherlands) and the ion semiconductor systems (Thermo Fisher Scientific) [12,13,16]. In recent years, most studies reporting NGS of bacterial pathogens have used Illumina short read sequencers. In contrast, third generation DNA sequencers yield much longer sequencing reads (i.e. > 10 kb) and can sequence single DNA molecules without the requirement for DNA amplification. Examples of third generation sequencing platforms include the Single Molecular Real-Time (SMRT) sequencing platforms manufactured by PacBio (Menlo Park, CA, USA) and the nanopore MinION sequencers manufactured by Oxford Nanopore Technologies (Oxford UK) [12]. The error rates generated by third generation sequencers are relatively high compared with second-generation sequencers, but the long-read lengths generated by the former are very advantageous for sequencing entire genomes and especially for plasmids. However, PacBio platforms are currently relatively expensive and are not suitable for the high-throughput rapid analysis or the processing of many clinical isolates. In contrast, Oxford Nanopore sequencers are inexpensive and can yield very long sequence reads, but bioinformatics analysis of sequence data is more challenging. Nonetheless, a growing range of nanopore bioinformatics tools are being developed, which soon will make this technology more readily applicable to clinical isolates [17]. A combination of short read sequences generated by NGS platforms such as Illumina MiSeq and long read sequences generated by nanopore sequencers can be used to rapidly and

accurately map relatively large genomic regions such as the staphylococcal cassette chromosome *mec* (SCC*mec*) region and other large genomic elements such as the arginine catabolic mobile element (ACME) in *S. aureus* and coagulase-negative staphylococcal species. Typically, the DNA sequences of such elements are spread across several contiguous sequences in genome assemblies generated from short read sequences and usually require extensive PCRs to accurately refine the assemblies. This can be overcome by generating hybrid assemblies of both short read and long read sequences [18,19]. A number of recent reviews have provided comprehensive overviews of whole genome sequencing (WGS) platforms and technology and its applications for microbial epidemiology [12-15].

Genome assembly and bioinformatics

The volume and complexity of data generated by WGS platforms requires the application of a variety of bioinformatics tools to determine the quality of the sequence data and to transform unrefined sequencing read data into more useful or meaningful forms. Software algorithms are used to clean up and organise sequence data, to assemble genome sequences from overlapping sequence reads, and to identify genomic variants, for genotyping and for phylogenetic analysis. For the non-bioinformatician, the terminology and application of bioinformatics is daunting in this rapidly developing field. A recent review by Carriço *et al.* (2018) provides an excellent overview of bioinformatics as applied to WGS data for the non-expert [20]. Some of the more frequently used bioinformatics approaches and software for microbial genome assembly and subsequent analysis have been recently reviewed [20,21,12]. Many software packages used for the analysis of WGS data are freely available and several commercially available and easy to use

142 software packages including BioNumerics (Applied Maths, Ghent, Belgium) and SeqSphere
143 (Ridom GmBH, Münster, Germany) are widely used for this purpose [12,14,16]

144 **Applications of WGS data for epidemiology**

145 Whole genome sequencing data can be used for a variety of purposes in investigating outbreaks
146 of infection and in tracking the sources and spread of infection in hospitals as well as in
147 investigating more regional and global aspects of the emergence of specific clones of particular
148 pathogens. Conventional targeted molecular typing of pathogens (e.g. PFGE or conventional
149 MLST) generates a genotype barcode or molecular fingerprint of each isolate based on a
150 relatively small portion of the genome [7, 12, 14]. Isolates with identical or similar genotypes
151 linked by epidemiological data are presumed to represent linked cases of infection. In contrast,
152 WGS enables the entire genome of isolates to be compared, which significantly enhances
153 resolution. As with conventional molecular typing, the genomes of isolates recovered from an
154 outbreak or cluster of infections are likely to be closely related.

155

156 *Single nucleotide variation analysis*

157 All microorganisms accumulate changes in their genomes over time, often through random
158 mutations. These include alterations to single nucleotide bases (single nucleotide variations or
159 SNVs). When an SNV becomes fixed within a population it is referred to as a single nucleotide
160 polymorphism or SNP. Other types of changes to the genome include gene acquisition by
161 horizontal transfer (e.g. plasmid acquisition or lysogenisation with a bacteriophage), small
162 insertions and deletions, gene duplication and genome rearrangements.

163

164 SNV analysis is frequently used to type isolates based on WGS data in an outbreak scenario and

has been used extensively using data generated by short read sequencers such as those manufactured by Illumina [12-15, 22-24]. This approach involves mapping sequence reads or larger contiguous sequences (known as contigs) assembled from overlapping short read sequences to a reference genome by core genome alignment. Specific software filtering tools can be used to exclude sequence stretches of high variability. Curated collections of reference genomes for particular species (including *S. aureus*) are available for SNV analysis. However, SNV analysis yields the best results when a reference genome that is closely related to the samples under investigation is used. In the case of an outbreak, isolates are likely to be closely related and one isolate sequence can be used as a reference against which other outbreak isolates' sequences can be mapped. The alignment of isolate sequences can then be used for phylogenetic analysis to determine the relationships between the isolates based on the identification of bases that differ in the test samples relative to the corresponding bases in the reference genome.

Extended multilocus sequence typing

The genetic relatedness of isolates can also be investigated using extended versions of conventional MLST. Conventional MLST typically involves the sequencing of segments of a small number of selected housekeeping genes that accumulate genetic changes relatively slowly because the encoded proteins are functionally constrained [25]. Such genes encode segments that are amplified by PCR using specific primers and the amplimers sequenced by conventional Sanger sequencing. Combinations of allelic variants in the selected genes are used to define a sequence type (ST) for a particular microorganism. The advent of WGS has enabled the establishment of whole genome (wg) and core genome (cg) MLST schemes that enable a comparison of test sequences with large curated sets of predefined genes for a particular species

[26,27]. Publicly available curated gene sets can include hundreds to thousands of genes depending on the scheme and analysis can be undertaken with easy to use software packages such as BioNumerics and SeqSphere [14]. The use of curated cgMLST schemes facilitates good interlaboratory reproducibility.

Thresholds of isolate relatedness

Establishing SNV and cgMLST thresholds of relatedness for a microorganism can be problematic. Genomic variability increases over time and it is vital that this consideration is borne in mind when attempting to assess the relatedness of isolates based on WGS data. There are no definitive rules for determining isolate relatedness or what constitutes a significant difference between isolates. Judgements on similarity or significantly different should be made separately according to the facts of each case. Meaningful thresholds of relatedness can be developed by investigating epidemiologically linked and unrelated isolates. However, proposed thresholds should always be interpreted in conjunction with epidemiological data.

A recent study that investigated several outbreaks of *S. aureus* using WGS suggested that in an acute short-term outbreak, there will be insufficient time for diversity to accumulate [23] and therefore establishing thresholds in this scenario should be less problematic. Schürch et al. (2018) recently detailed a list of current suggested SNV and cgMLST relatedness criteria for some representative clinically significant bacterial species [14]. It is worth bearing in mind that clonality thresholds may vary within particular clones of a particular pathogen.

The use of WGS for investigating MRSA outbreaks

The more widespread availability of WGS in recent years has enabled us to study in more detail patterns of spread, including detecting previously undocumented transmission, as well as the overall and detailed evolution of strains of MRSA [28,29]. A study of Danish isolates using *spa* typing and Sanger sequencing found a 97% agreement between both these methods, and WGS [30]. As the technology becomes increasingly available and costs reduce, WGS will be no longer confined to research or reference laboratory facilities but become increasingly more available in routine clinical laboratories to inform infection prevention and control strategies, as well as outbreak management in real time [30-32].

Price and colleagues in 2013 outlined the potential of WGS, discussed some recent applications and highlighted its potential for the future [33]. We undertook to update that and highlight important findings relating to the detection of outbreaks, their evolution over time, and inter-hospital spread, and how this could potentially benefit preventative measures for MRSA in the community. We did not set out to cover aspects of WGS and MRSA as they relate specifically to antimicrobial resistance and global molecular epidemiology. A literature search was undertaken of articles in PubMed, Embase, the Cochrane Library and Web of Science for articles up to the end of August 2018. Search terms included MRSA, WGS, outbreaks, clinical, infection, prevention and control. In total, 588 items were found, and with duplicates removed, and having reviewed all titles and abstracts for relevance (e.g. excluding those related purely to the veterinary setting), those cited focus on the value and relevance of WGS to MRSA in terms of prevention and control and outbreak investigation. We excluded studies where the focus was purely on clonal evolutionary trends and were not germane to infection prevention and control practitioners. What follows is a discussion of the transmission of MRSA and especially outbreak

management. Studies that refer to and or include methicillin-susceptible *S. aureus* (MSSA) are included were MRSA isolates were also included, and or, where the findings might equally apply to outbreaks or settings involving MSSA.

Young and colleagues have looked at the evolutionary dynamics of *S. aureus* from carriage to disease, i.e. a patient who carried *S. aureus* in the nose and who subsequently developed bloodstream infection (BSI) [34]. Just eight mutations accompanied the transmission from carriage to infection [34]. However, during an outbreak which occurs over a matter of weeks or even months, much of the literature seems to agree that up to approximately 30 SNPs may be allowed between isolates before the isolates are considered different, i.e. transmission may have occurred between two patients if their isolates differ by less than 30 SNPs [11, 14]. Such conclusions are predicated on the assumption that the epidemiological findings are supportive.

Epidemiology and control

Neonatal and paediatric units

The occurrence of MRSA amongst neonates has potentially devastating consequences. Hence, many studies of WGS and MRSA have focussed on neonates or paediatric units, because of the serious clinical consequences. Whole genome sequencing has been used to track the spread of MRSA and to assist in early intervention measures [26, 35-43].

Köser and colleagues investigated a putative MRSA outbreak in the UK and WGS revealed a distinct cluster with clear separation between outbreak and non-outbreak isolates, amongst a collection of ST22 isolates [37]. When using less-discriminatory methods of typing, the extent of an outbreak may be exaggerated. In a neonatal unit outbreak involving 17 neonates, *spa* typing,

PFGE and WGS were used. All 17 isolates belonged USA-300 isolates according to PFGE. However, while five isolates were involved in recent transmission events, 12 (70.5%) represented genetically unique isolates according to WGS and were therefore believed not to be part of the outbreak [37]. This finding is important as it suggests that there was no obvious deficiency in infection prevention measures by healthcare staff because there may be have been multiple independent introductions of USA-300.

Earls and colleagues investigated two protracted outbreaks (2009-2011 and 2014-2017) in the neonatal intensive care unit (NICU) of an Irish hospital involving clonal complex (CC) 88-MRSA isolates belonging to *spa* types t186 and t786 [42]. Isolates were recovered from 20 separate neonates during the outbreaks, together with two isolates recovered two years apart from the same healthcare worker. Whole genome sequencing and subsequent wgMLST analysis revealed that both outbreaks were caused by the same CC88/ST78-MRSA-IVa strain. All the isolates formed a large cluster, exhibiting 1–71 pairwise allelic differences in a wgMLST-based minimum spanning tree (MST). The maximum distance observed between any two directly linked nodes was 32 alleles, detected between two t186 isolates, which were recovered almost three years apart during outbreak one. All other directly linked isolates exhibited 1–19 allelic differences [42]. This indicated a high degree of relatedness between all isolates within the cluster network. There were no apparent sub-clusters based on *spa* type and the one direct link within the MST between *spa* type t786 isolates and *spa* type t186 isolates corresponded to an allelic difference of 18. The two t786 isolates recovered from a healthcare worker two years apart exhibited 20 allelic differences and differed from other t786 isolates by 10-21 and 9-37 allelic differences, respectively, indicating the involvement of the HCW in the outbreak

transmission. Unfortunately, no information on whether the HCW was persistently or transiently colonised with the CC88/ST78-MRSA-IVa strain during the two year period was available or if attempts to decolonise the HCW were undertaken. This study also demonstrated the spread of the ST78-MRSA-IVa strain to two other Irish hospitals. A cgMLST-based comparison with international comparator isolates showed that the outbreak strain was most likely imported from Australia, where it is among the prevalent MRSA clones.

The French national staphylococcal reference laboratory used WGS to retrospectively investigate MRSA isolates amongst four separate clonal complexes (CC1, CC5, CC8, CC30) involved in community and hospital outbreaks, which included 41 CC5 isolates from new-borns [43]. Even though the *spa* type was different, isolates from Limoges and Bordeaux, which are separated by approximately 180 km, differed by less than 22 SNPs suggesting that despite the geographical difference, they could be part of the same transmission pattern [43]. The value of WGS in this setting, are outlined in Table 2.

Other hospital outbreaks

Studies have shown the value of WGS in teasing out some of the subtleties of general and hospital outbreak evolution and development [11,16,23,44-50]. The isolates from three separate outbreaks were studied, including one in a hospital, resulting in a total of 42 isolates; 15 of 16 isolates from a burns unit formed a single cluster but 12 isolates from a post-surgical unit were more diverse; the authors concluded that those with less than eight SNPs should be considered related, and those between nine to 29, as being possibly unrelated [16].

We have assessed the variability amongst MRSA isolates collected from patients and the environment in a prospective study that involved 41 patient and environmental isolates that were sequence type 22 [11]. We traced the isolates in terms of the geographical location and the time when they were recovered. Far more combinations of isolates, i.e. patient-patient or patient-environment, indicating potential transmission links, were detected by WGS compared with conventional molecular typing using *spa*, *dru* and or PFGE typing or a combination of all three [11]. The *dru* region is a noncoding DNA segment consisting of imperfect 40-bp variable-number tandem repeats (VNTRs) located in the hypervariable region between *mecA* and IS431*mec* of SCC*mec* [7,11].

In a study on a Dutch oncology ward involving an asymptomatic nasal colonized healthcare worker, WGS MLST showed similarities between MSSA and MRSA isolates involved in an outbreak and the authors hypothesised that a fusidic acid resistant isolate of MSSA acquired a SCC*mec*, and subsequently caused an MRSA outbreak suggested a genetic link [46]. Miller and colleagues were concerned about the patterns of MRSA BSIs in a specific hospital in England, involving a clonal variant of EMRSA-16. Isolates causing BSI between 2000 and 2001, and 2006 to 2007 were investigated [48]. The clonal variant was largely confined to that hospital unlike isolates causing BSIs acquired elsewhere, and infections caused by this clone were significantly associated with increased peripheral white cell and neutrophil count, suggesting increased virulence [48]. Similarly, a retrospective review of isolates in 2004-2014 was undertaken in Switzerland to reconstruct transmission pathways [49]. Tracking the geographic locations of patients who were colonised or infected together with WGS data, enabled the researchers to assess patterns of spread, which included a network of hospitals and overlapping periods of

hospitalisation. In one case, an outbreak lasted several months in an orthopaedic ward, but was only retrospectively detected using WGS. Looking at isolates from the same patient over time, the authors concluded that there was one SNP every 8.9 weeks or 0.016 per day [50].

Tong *et al*. (2015) used WGS to investigate the genetic diversity of ST239 MRSA isolates from patients over a three-month period in two ICU units of a 1000 bed hospital in Thailand where transmission was common [22]. Phylogenetic analysis revealed a flux of distinct ST239 clades (or groups of isolates) over time in each ICU. Analysis of WGS data confirmed intra-ward and inter-ward transmission events and revealed that one patient in each ICU was the source of numerous transmission events. The mean pairwise SNP differences between the five ST239 clades identified was ≥ 197 SNPs, indicating that each clade was distinctly different.

Community and wider patterns of spread

While the focus of the hospital infection prevention and control team is to largely prevent and analyse outbreaks within the hospital, these may sometimes arise from outside the hospital, i.e. from patients admitted from other hospitals or from LSRU as well as potentially from patients who have been abroad, especially if hospitalised there. Hence, the analysis of a wider range of isolates can inform preventative strategies and highlight the innate capacity of *S. aureus* to spread and evolve.

The ST8 USA300 MRSA clone emerged shortly after 2000 and subsequently became the leading cause of skin and soft tissue infections in the United States (US). The origin of USA300 in Pennsylvania region of the US and its subsequent range expansion was recently investigated in

351 detail using genome sequences from 357 isolates from 22 states and territories and seven other
352 countries [51]. USA300 is now common internationally. Fluit and colleagues compared one
353 well-characterised strain from the US with those from Europe [52]. There was a difference of
354 144 SNPs between the US isolate and those from Europe, the gene content showed 21 regions of
355 difference, and the European strains were resistant to fewer antibiotics. However, the SNP data
356 suggested a common ancestor around two decades ago [52]. This clone is an important pathogen
357 internationally, even if not as common in Europe as in the US, but that could change with spread
358 via ongoing international travel.
359
360 An outbreak of ST97-IVa involving 25 patients, originating from a surgical ward, over a four-
361 year period in Denmark, a country with a low prevalence of MRSA, was investigated by WGS
362 [53]. Eighteen patients had been admitted to the surgical ward of which 13 overlapped in terms
363 of admission periods. Two HCWs and two patient family members were also involved in the
364 outbreak. All except two isolates were *spa* type t267 and belonged to ST97. In this outbreak,
365 WGS linked nine initial isolates to 16 previous isolates, resulting in 23 patients being involved
366 with the suggestion that a healthcare worker with undetected carriage may have caused the
367 outbreak [53]. The authors noted that with shorter lengths of stay, patients may not be identified
368 as being colonized with MRSA carriers while in hospital but on follow up in the community after
369 discharge from hospital [53].
370
371 Modern healthcare requires patients to be transferred between hospitals as specialist and tertiary
372 facilities are centralised. A comparison of EMRSA-15 within the UK and Ireland using WGS
373 showed that the hospitals within the same referral regions had similar MRSA populations but

transmission within a hospital arose from patients having been transferred from another hospital [54]. Furthermore, frequent patient admissions to multiple hospitals results in ward-based transmission within a hospital, as detected by a study in two NHS hospital groups and a district general hospital in South-East London, involving ST22 MRSA isolates [55].

As the age of the population increases, patients are discharged more quickly than before from acute hospitals to LSRU, such as to nursing homes. Sometimes these patients require re-admission and there is constant flux between the acute and long-stay sector. Many studies have highlighted this dynamic in terms of the acquisition and transmission of MRSA between these sectors [56-62]. Furthermore, there has been interest in the role of carriage, including enteric carriage [44,50]. In a long-term outbreak involving 1,600 patients and where WGS was used to determine the origin, a single clonal variant of ST228 was responsible but that this clone was more frequently recovered from the groin and rectal swabs [50].

In Singapore, where 1700 hospital patients and LSRU occupants were screened for MRSA over a 6-week period, MRSA prevalence was lowest in acute healthcare facilities, i.e. 11.8% compared with intermediate or long-term care facilities, 29.95 and 20.4%, respectively [56]. Furthermore, LSRUs had the greatest diversity of MRSA clones. Stine and colleagues examined the transmission of MRSA from resident-to-resident in LSRUs using WGS. Multiple sites from residents were screened over a 12-week study [59]. Isolates from multiple body sites were usually closely related and many residents living together often harboured closely related strains [59].

Following a protracted hospital outbreak between June 2013 and June 2016 caused by multidrug resistant ST1-MRSA-IV isolates belonging to *spa* type t157, a collection of 89 isolates from the outbreak hospital, 16 other hospitals and four other healthcare facilities and the community in Ireland were investigated by WGS [61]. Fifty of the isolates including 40 from the outbreak exhibited high-level mupirocin resistance mediated by a *iles2*-encoding plasmid conjugative plasmid [61]. Pairwise SNVs exhibited by healthcare-associated and community-associated isolates indicated recent transmission of ST1-MRSA-IV within and between multiple hospitals, healthcare facilities and communities in Ireland [60]. This has implications for current MRSA prevention and control guidelines, which are very much focused on measures in the acute hospital sector. More recent studies have identified the multidrug resistant ST1-MRSA-IV clone as a novel CC1-MRSA-IV clone that has recently emerged in several European countries (see section on the identification of emerging MRSA clones below).

When developing guidelines, the question of what measures to take within households where there is an MRSA-positive individual, often arises. Current Irish recommendations are to highlight personal standards of hygiene and cleanliness and to minimise disruption/inconvenience in a home, as intra-familial spread is not considered common [63]. A retrospective study of isolates collected between 2008 and 2010 in Chicago and Los Angeles found very little genetic variation amongst USA300 isolates within households but that transmission did occur where the index patient had skin and soft tissue infection [64]. In England, where USA300 is less common than in the United States, a 12-month prospective observational study of 2283 screening and clinical isolates from 1465 patients collected between April 2012 and April 2013 from the community and hospitals, found USA300 in only 24 cases

(1.6%). There were also three groups of closely related isolates with a maximum genetic distance of 6, 59 and 9 SNPs, respectively, amongst epidemiologically linked cases [65]. The authors concluded that international travel may have played a role in the introduction of this clone into England.

Over the last decade, it has been increasingly recognised that some patients without a recent history of healthcare contact who develop MRSA may have done so via contact with livestock, i.e. livestock-associated MRSA (LAMRSA). This has been particularly well described in those countries with a low background prevalence of hospital MRSA such as Denmark. Larsen and colleagues analysed human cases of LA-MRSA during 2010-2015 in Denmark [66]. Seventeen cases of BSI, 700 cases of skin and soft tissue infection and 76 cases of other infections due to LAMRSA, were studied. Overall, 32% of the LAMRSA were from individuals with no contact with livestock. Whole-genomes sequence analysis suggested that most isolates were closely related to Danish pig isolates [66]. A study of CC398, commonly LAMRSA, was assessed in human and pig isolates in Norway [67]. A human case was identified in 2009 but by the end of 2014, there were a total of 84 human cases. Epidemiological links placed these individuals in three clusters and all farms had farm workers originating from other European countries where MRSA is more prevalent [67]. Interestingly, while the farm workers and other possible human carriers may have been non-Norwegian, none of the farms had imported pigs from abroad, and the transmission of this clonal complex may have been by human introduction via migration rather than through the importation of pigs. The results of this study would seem to justify the Norwegian control strategy of targeting the screening of personnel before working in pig herds

as part of national surveillance [65]. Contact with livestock is therefore a potential risk factor to be considered and not just in countries with a low prevalence of MRSA in acute hospitals.

National surveillance systems together with key performance indicators increasingly drive reductions in serious infections due to MRSA, particularly BSI. In the UK, a "zero tolerance" approach has been advocated but there may be a portion of MRSA BSIs that are not preventable. In one instance, WGS assisted in determining if an outbreak was preventable through infection prevention and control measures. A study in Cambridge, UK, focussed on a cluster of five MRSA BSIs between September 2011 and August 2012. The researchers used detailed epidemiological methods and WGS analysis of isolates and found that there were varying degrees of overlap in admission to the wards of these cases [68]. A comparison of isolates indicated that each patient was infected by their own carriage isolate. Amongst four of the patients, isolates differed by between 122 and 168 cgSNPs [68]. A wider analysis of all patients with MRSA, whether colonised or infected was undertaken including the cases of MRSA BSIs. Three of the five BSI episodes were associated with skin conditions and two were attributable to intravascular catheters. From an analysis of the WGS data, it was concluded that these cases were not due to inter-patient transmission [68]. In this setting, WGS suggested that there was no deficiency of infection prevention and control services in preventing these five cases of BSI, and that some or all these cases of BSI were probably not preventable.

Identification of emerging MRSA clones

A study by Earls *et al*. (2019) used WGS to investigate the recent emergence of multidrug-resistant Panton-Valentine leukocidin (PVL)-negative CC1-MRSA-IV isolates in multiple Irish

hospitals and the community and in two hospitals in the German City of Regensburg between 2016-2018 where it was also identified in the community [69**]**. Phylogenetic analysis grouped the isolates into a large clade, where no isolate differed from any other isolate by more than 130 cgSNVs. Clade isolates harboured an SCC*mec* type IVa element with a characteristic 4710 nucleotide insertion in the downstream constant segment (*dcs*) adjacent to *orfX* and harboured the same allelic variants of the SCC*mec* genes, *ccrA2* (1350 bp) and *ccrB2* (1629 bp). Overall, clade isolates exhibited genotypic characteristics which differed comprehensively from those associated with other previously well characterised CC1-MRSA-IV clones including Western Australia (WA) MRSA-1 and USA400. Five MRSA isolates recovered in a Romanian hospital between 2010 and 2012 and 10 CC1 MSSA isolates recovered in the same Romanian hospital between 2009 and 2012 also grouped into the novel CC1 clade. Earls *et al*. (2019) designated the multidrug resistant MRSA isolates as a novel European clade of CC1-MRSA-IV and hypothesized that this clade likely recently emerged from CC1 MSSA in Romania or a neighbouring country **[69].** Interestingly, isolates of this emerging European CC1-MRSA-IV clone were also recently identified in an Italian paediatric hospital [24].

Routine applications

Having data interpreted within two to three days or less, will enhance infection prevention and control measures and may prevent the use of disruptive measures such as the closing of units or wards. Eyre and colleagues used rapid benchtop sequencing to investigate two outbreaks of *S. aureus* within five days of a positive culture result [70]. One involved ten MRSA isolates from eight patients in an intensive care unit and the other involved six patients over three months in the south of England with Panton-Valentine leucocidin positive MRSA. Within clusters or

outbreaks, most isolates were indistinguishable, and all were within three SNVs [70]. On a broader public health front, WGS may be used to track community and hospital-acquired isolates within and between countries. A total of 308 invasive isolates collected across Europe identified predominant clones, e.g. clonal complexes 5 and 22, with the latter representing EMRSA-15 originally from the UK [71].

Conclusions

The application of WGS has greatly expanded our knowledge of clinical and epidemiologic aspects of MRSA infection and colonization, including transmission, the commonality of clones in the community as well as the evolution of resistant determinants. Subsequent studies will further inform our understanding of transmission patterns and guide what interventions are most important and when they should be applied. The use of WGS in day-to-day practice will be facilitated by improvements in our capacity to interpret the data and apply it appropriately, in a timely manner. This may in many instances confirm low level but ongoing clusters and cross-transmission of MRSA, which if acted on, may assist in preventing larger outbreaks. The availability of WGS data may also suggest considerable patient-to-patient transmission of a single clone or that an outbreak is due to the simultaneous emergence of different clones even if patients are temporally and geographically linked. Alternatively, an outbreak may arise from MRSA being carried into a hospital, from either other hospitals or a LSRU.

Future priorities include agreeing parameters for deciding whether isolates are closely related or otherwise enhanced data analysis and ensuring IT systems can share WGS data between hospitals, between countries and further afield. As this field continues to mature and evolve, WGS may potentially have a significant role in informing measures to prevent transmission through the provision of critically important molecular epidemiological information in real time.

517 ***Declaration:*** HH has been in receipt of research funding from Pfizer and Astellas and has
518 provided educational sessions or advice to Pfizer and Cepheid in recent years.

519 ***Funding:*** The preparation and drafting of this review was not supported by any external funding.
520 However, the authors would like to acknowledge recent research funding from the Irish Health
521 Research Board (grant number HRA-POR-2015-1051) and previously from Scientific
522 Foundation Ireland for research on MRSA in recent years. We also acknowledge and thank Paul
523 J Murphy for assistance in the literature search.

524

References

1. Zingg W, Hopkins S, Gayet-Ageron A, Holmes A, Sharland M, Suetens C for the ECDC PPS study group. Health-care-associated infections in neonates, children, and adolescents: an analysis of paediatric data from the European Centre for Disease Prevention and Control point-prevalence survey. Lancet Infect Dis 2017; 17: 381-89.
2. Smyth ETM, McIlvenny G, Enstone JE, Emmerson AM, Humphreys H, Fitzpatrick F, *et al*. Four country healthcare associated infection prevalence survey 2006: overview of the results. J Hosp Infect 2008; 69: 230-248.
3. Guy R, Geoghegan L, Heginbothom M, Howe R, Muller-Pebody B, Reilly JS, *et al*. Non-susceptibility of *Eschericihia coli*., *Klebsiella* spp., *Pseudomonas* spp., *Streptococcus pneumoniae* and *Staphylococcus aureus* in the UK: temporal trends in England, Northern Ireland, Scotland and Wales. J Antimicrob Chemother 2016; 71: 1564-1569.
4. Mizuno S, Iwami M, Kunisawa S, Naylor N, Yamashita K, Kyratsis Y, *et al.* Comparison of national strategies to reduce meticillin-resistant *Staphylococcus aureus* infections in Japan and England. J Hosp Infect 2018; 100: 280-298.
5. Monecke S, Slickers P Ehricht R. Assignment of *Staphylococcus aureus* isolates to clonal complexes based on microarray analysis and pattern recognition. FEMS Immunol Med Microbiol 2008; 53: 237–51.
6. Monecke S, Jatzwauk L, Müller E, Nitschke H, Pfohl K, Slickers P, et al. Diversity of SCC*mec* elements in *Staphylococcus aureus* as observed in south-eastern Germany. PLoS One 2016; 11(9): e0162654. doi: 10.1371/journal.pone.0162654.
7. Shore AC, Rossney AS, Kinnevey PM, Brennan OM, Creamer E, Sherlock O, Dolan A, *et al*. Enhanced discrimination of highly clonal ST22-methicillin-resistant *Staphylococcus aureus* IV isolates achieved by combining *spa*, *dru* and pulsed-field gel electrophoresis typing data. J Clin Microbiol 2010; 48:1839–52.
8. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, *et al*. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. Genome Res 2013; 23:653–64.
9. Grundmann H, Schouls LM, Aanensen DM, Pluister GN, Tami A, Chlebowicz M, *et al*. ESCMID Study Group on Molecular Epidemiological Markers, European Staphylococcal Reference Laboratory Working Group. The dynamic changes of dominant clones of *Staphylococcus aureus* causing bloodstream infections in the European region: results of a second structured survey. Euro Surveill 2014; 19(49):pii_20987. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId_20987.
10. Creamer E, Shore AC, Rossney AS, Dolan A, Sherlock O, Fitzgerald-Hughes D, et al. Transmission of endemic ST22-MRSA-IV on four acute hospital wards investigated using a combination of *spa*, *dru*, and pulsed-field gel electrophoresis typing. Eur J Clin Microbiol Infect Dis 2012; 31: 3151-3161.
11. Kinnevey PM, Shore AC, MacAogáin M, Creamer E, Brennan GI, Humphreys H, *et al*. Enhanced tracking of nosocomial transmission of endemic sequence Type 22 methicillin-resistant *Staphylococcus aureus* type IV isolates among patients and environmental sites by use of whole-genome sequencing. J Clin Microbiol 2016; 54: 445-448.
12. Tang P, Croxen MA, Hasan MR, Hsiao WW, Hoang LM. Infection control in the new age of genomic epidemiology. Am J Infect Control 2017; 45:170–9.

13. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 2018; 24:335-41.
14. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect 2018; 24:350–4.
15. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. Microbiol 2018; 164:1213–9
16. Cunningham SA, Chia N, Jeraldo PR, Quest DJ, Johnson JA, Boxrud DJ, et al. Comparison of whole-genome sequencing methods for analysis of three methicillin-resistant *Staphylococcus aureus* outbreaks. J Clin Microbiol 2017; 55: 1946-1953
17. Leggett RM, Clark MD. A world of opportunities with nanopore sequencing. J Exp Bot 2017; 68:5419-29.
18. Lemon JK, Khil PP, Frank KM, Dekker JP. Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. J Clin Microbiol 2017; 55:3530-43.
19. Baig S, Johannesen TB, Overballe-Petersen S, Larsen J, Larsen AR, Stegger M. Novel SCC*mec* type XIII (9A) identified in an ST152 methicillin-resistant *Staphylococcus aureus*. Infect Genet Evol 2018; 61:74-6.
20. Carriço JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. Clin Microbiol Infect 2018; 24:342–9.
21. Oakeson KF, Wagner JM, Mendenhall M, Rohrwasser A, Atkinson-Dunn R. Bioinformatic analyses of whole-genome sequence data in a public health laboratory. Emer Infect Dis 2017; 23:1441-5.
22. Tong SY, Holden MT, Nickerson EK, Cooper BS, Köser CU, Cori A, *et al*. Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. Genome Res 2015; 25:111–8.
23. Gordon NC, Pichon B, Golubchik T, Wilson DJ, Paul J, Blanc DS, *et al*. Whole-genome sequencing reveals the contribution of long-term carriers in *Staphylococcus aureus* outbreak investigation. J Clin Microbiol 2017; 55:2188–97.
24. Manara S, Pasolli E, Dolce D, Ravenni N, Campana S, Armanini F, *et al*. Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital. Genome Med 2018; 10(1):82 doi: 10.1186/s13073-018-0593-7.
25. Enright MC, Day NPJ, Davies CE, Peacok SJ, Spratt BG. Sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol 2000;38: 1008-1015.
26. Leopold SR, Goering RV, Witten A Harmsen D, Mellmann A. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol* 2014, 52, 2365–70.
27. Roisin S, Gaudin C, De Mendonça R, Bellon J, Van Vaerenbergh K, De Bruyne K. *et al*. Pan-genome multilocus sequence typing and outbreak-specific reference-based single nucleotide polymorphism analysis to resolve two concurrent *Staphylococcus aureus* outbreaks in neonatal services. Clin Microbiol Infect 2016; 22:520–6.
28. Gilchrist CA, Turner SD, Riley MF, Petri Jr WA, Hewlett EL. Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev 2015; 28: 541-563.

29. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. Curr Opin Microbiol. 2015; 23:62-7.
30. Bartels MD, Petersen A, Worning P, Nielsen JB, Larner-Svensson H, Johansen HK, *et al.* Comparing whole-genome sequencing with Sanger sequencing for *spa* typing of methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol 2014; 52: 4305-4308.
31. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 2017; 30:1015-1063.
32. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N *et al*. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010; 327: 469-474.
33. Price JR, Didelot X, Crook DW, Llewelyn MJ, Paul J. Whole genome sequencing in the prevention and control of *Staphylococcsu aureus* infection. J Hosp Infect 2013; 83: 14-21.
34. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, *et al.* Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. PNAS 2012; 109: 4550-4555.
35. Nübel U, Nachtnebel M, Falkenhorst G, *et al.* MRSA Transmission on a neonatal intensive care unit: epidemiological and genome-based phylogenetic analyses. *PLoS* 2013; 8(1): e54898. doi: 10.137/jopurnal.pone.0044898
36. Layer F, Sanchini A, Strommenger B, Cuny C, Breier A-C, Proquitté H, *et al*. Molecular typing of toxic shock syndrome toxin-1- and enterotoxin A-producing methicillin-sensitive *Staphylococcus aureus* isolates from an outbreak in a neonatal intensive care unit. Internat J Med Micro 2015; 305: 790-798.
37. Köser CU, Holden MTG, Ellington MJ,Ellington MJ, Cartwright EJP, Brown NM, *et al*. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl Med 2012; 366: 2267-75.
38. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, *et al.* Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*; a descriptive study. Lancet Infect Dis 2013; 13: 130-136.
39. Azaraian T, Maraqa NF, Cook RL, Johnson JA, Bailey C, Wheeler S, *et al.* Genomic epidemiology of methicillin-resistant *Staphylococcus aureus* in a neonatal intensive care unit. PLoS One 2016; 11(10): e0164397. doi:10.371/journal.pone.0164397.
40. Azarian T, Cook RI, Johnson JA, Guzman N, McCarter YS, Gomez N, *et al*. Whole-genome sequencing for outbreak investigations of methicillin-resistant *Staphylococcus aureus* in the neonatal intensive care unit: time for routine practice? Infect Control Hosp Epidemiol 2015; 36(7): 777-785.
41. Ugolotti E, Di Marco E, Bandettini R, Biassoni R. Genomic characterisation of a paediatric MRSA outbreak by next-generation sequencing. J Hosp Infect 2018; 98: 155-160.
42. Earls MR, Coleman DC, Brennan GI, Fleming T, Monecke S, Slickers P, *et al.* Intra-hospital, inter-hospital and intercontinental spread of ST78 MRSA from two neonatal intensive care units outbreaks established using whole genome sequencing. Front Microbiol 2018; 9.1485.doi: 10.3389/fmicb.2018.01485.
43. Durand G, Javerliat F, Bes M, Veyrieras J-B, Guigon G, Mugnier N, *et al.* Routine whole-genome sequencing for outbreak investigations of *Staphylococcus aureus* in a national reference center. Front Microbiol 2018; 9.511. doi: 10.3389/fmicb.2018.00511.

44. Kong Z, Zhao P, Liu H, Yu X, Qin Y, Su Z, *et al.* Whole-genome sequencing for the investigation of a hospital outbreak of MRSA in China. PLoS ONE 2016; 11(3): e0149844, doi: 10.1371/journal.pone.0149844.
45. Price JR, Cole K, Bexley A, Kostiou V, Eyre DW, Golubchik T, *et al*. Transmission of *Staphylococcus aureus* between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. Lancet Infect Dis 2017; 17: 207-14.
46. Weterings V, Bosch T, Witteveen S, Landman F, Schouls L, Klutyman J. Next-generation sequence analysis reveals transfer of methicillin resistance to a methicillin-susceptible *Staphylococcus aureus* strain that subsequently caused a methicillin-resistant *Staphylococcus aureus* outbreak: A descriptive study. J Clin Microbiol 2017; 55: 2808-2816.
47. Long WS, Beres S, Olsen RJ, Musser JM. Absence of patient-to-patient intra-hospital transmission of *Staphylococcus aureus* as determined by whole-genome sequencing. mBio 2014; 5(5): e01692-14.doi.10.1128/mbio.01692-14.
48. Miller RR, Price JR, Batty EM *et al.* Healthcare-associated outbreak of methicillin-resistant *Staphylococcus aureus* bacteraemia: role of a cryptic variant of an epidemic clone. *J Hosp Infect* 2014; 86: 83-89.
49. Moore G, Cookson B, Gordon NC, Jackson R, Kearns A, Singleton J, *et al*. Whole-genome sequencing in hierarchy with pulsed-field gel electrophoresis: the utility of this approach to establish possible sources of MRSA cross-transmission. J Hosp Infect 2015; 90: 38-45.
50. Senn L, Clerc O, Zanetti G, Bassett P, Prod'hom G, Gordon NC, *et al*. The stealthy superbug: The role of asymptomatic enteric carriage in maintaining a long-term hospital outbreak of ST228 methicillin-resistant *Staphylococcus aureus.* mBio 2016; 7(1): e02039-15.doi.10.1128/mBio.02039-15.
51. Challagundla L, Luo X, Tickler IA, Didelot X, Coleman DC, Shore AC, *et al*. Range expansion and the origin of USA300 North American epidemic methicillin-resistant *Staphylococcus aureus*. MBio 2018; 9(1):e02016-17. doi: 10.1128/mBio.02016-17.
52. Fluit, AC, Carpaij N, Majoor EAM, Weinstein RA, Aroutcheva A, Rice TW, *et al.* Comparison of an ST80 MRSA strain from the USA with European ST80 strains. J Antimicrob Chemother 2015; 70: 664-669.
53. Rubin IM, Hansen TA, Klingenberg AM, Petersen AM, Worning P, Westh H, *et al.* A sporadic four-year hospital outbreak of a ST97-IVa MRSA with half of the patients identified in the community. Front Microbiol 2018; 9.1494. doi: 10.3389/fmicb.2018.0194.
54. Donker, T, Reuter S, Scriberras J,Reynolds R, Brown NM, Török ME *et al*. Population genetic structuring of methicillin-resistant *Staphylococcus aureus* clone EMRSA-15 within UK reflects patient referral patterns. Microbial Genom 2017; 3: e000113. Doi.10.1099/mgen.0.000113.
55. Auguet, OT, Stabler RA, Betley J, Preston MD, Dhaliwal M, Ioannou A, *et al*. Frequent undetected methicillin-resistant *Staphylococcus aureus* ward-based; transmission linked to patient sharing between hospitals. Clin Infect Dis 2018; 66: 840-8.
56. Chow A, Lim VW, Khan A, Pettigrew K, Lye DCB, Kanagasabai K, *et al*. MRSA transmission dynamics among interconnected acute, intermediate-term, and long-term healthcare facilities in Singapore. Clin Infect Dis 2017: 64: S76-S81.
57. Toleman, MS, Watkins ER, Williams T, Blane B, Sadler B, Harrison EM *et al*. Investigation of a cluster of sequence type 22 methicillin-resistant *Staphylococcus aureus* transmission in a community setting. Clin Infect Dis 2017; 65: 2069-2077.

58. de la Gandara MP, Garay JAR, Mwangi M, Tobin JM, Tsang A, Khalida C, *et al*. Molecular types of methicillin-resistant *Staphylococcus aureus* and methicillin-sensitive *S. aureus* strains causing skin and soft tissue infections and nasal colonization, identified in community health centers in New York City. J Clin Microbiol 2015; 53: 2648-2658.
59. Stine, OC, Burrowes S, David S, Johnson JK, Roghmann M-C. Transmission clusters of methicillin-resistant *Staphylococcus aureus* in long-term care facilities based on whole-genome sequencing. Infect Control Hosp Epidemiol 2016*; 37*: 685-691.
60. Auguet TO, Betley JP, Stabler RA. Evidence for community transmission of community-associated but not health-care-associated methicillin-resistant *Staphylococcus aureus* strains linked to social and material deprivation: spatial analysis of cross-sectional data. PLoS Med 2016: 13. Doi.10.1371/journal.pmed.1001944.
61. Earls MR, Kinnevey PM, Brennan GI, Lazaris A, Skally M, O'Connell B, *et al*. The recent emergence in hospitals of multidrug-resistant community-associated sequence type 1 and *spa* type t127 methicillin-resistant *Staphylococcus aureus* investigated by whole-genome sequencing: Implications for screening. PLoS ONE 2017: 12; e0175542.foi.10.1371/jouranl.pone.0175542.
62. Harrison EM, Zudden C, Broderick HJ, Blane B, Brennan G, Morris D *et al*. Transmission of methicillin-resistant *Staphylococcus aureus* in long-term care facilities and their related healthcare networks. Genome Med 2016; 8: doi 10.1186/s13073-016-0353-5.
63. National Clinical Effectiveness Committee. Prevention and control of methicillin-resistant *Staphylococcus aureus*. National Clinical Guideline No.2. Department of Health, Dublin, Ireland, September 2013. http://www.hpsc.ie/a-z/microbiologyantimicrobialresistance/infectioncontrolandhai/guidelines/File,14478,en.pdf (accessed 26th January 2019).
64. Alam MT, Read TD, Petit III RA, Petit III RA, Boyle-Vavra S, Miller LG *et al.* Transmission and microevolution of USA300 MRSA in U.S. households: Evidence from whole-genome sequencing. mBio 2015; 6(2): e000-54-15. Doi.10.1128/mBio.00054-15.
65. Toleman MS, Reuter S, Coll F, Harrison EM, Blane B, Brown NM, *et al*. Systematic surveillance detects multiple silent introductions and household transmission of methicillin-resistant *Staphylococcus aureus* USA 300 in the East of England. J Infect Dis 2016; 214: 447-453.
66. Larsen J, Peterson A, Larsen AR, Sieber RN, Stegger M, Koch A, *et al*. Emergence of livestock-associated methicillin-resistant *Staphylococcus aureus* bloodstream Infections in Denmark. Clin Infect Dis 2017: 65: 1072-1076.
67. Grøntvedt, CA, Elstrøm P, Stegger M, Skov RL, Andersen PS, LArssen KW, *et al*. Methicillin-resistant *Staphylococcus aureus* CC398 in humans and pigs in Norway: A "One Health" perspective on introduction and transmission. Clin Infect Dis 2016; 63: 1431-1438.
68. Török, ME, Harris SR, Cartwright EJP, Raven KE, Brown NM, Allison MED, *et al*. Zero tolerance for healthcare-associated MRSA bacteraemia: is it realistic? J Antimicrob Chemother 2014; 69: 2238-2245.
69. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, *et al*. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJOpen 2012; 2: e001124.doi:10.1136/bmjopen-2012-0011124.

70. Earls MR, Shore AC, Brennan GI, Simbeck A, Brachert WS, Vremeră T, *et al*. A novel multi-resistant PVL-negative CC1-MRSA-IV clone emerging in Ireland and Germany likely evolved in South-Eastern Europe. Infect Genet Evol 2019; 69:117-126.
71. Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, *et al*. Whole-genome sequencing for routine pathogen surveillance in public health; a population snapshot of invasive *Staphylococcus aureus* in Europe. mBio 2016; 7(3): e00444-16.doi:10.1128/mbio.oo4444-16.

Table I. List of commonly used terms relating to whole genome sequencing and analysis

Term	Explanation
Next generation sequencing (NGS)	The high-throughput and rapid determination of whole genome sequences. Millions to billions of DNA fragments are sequenced in parallel.
Sequencing platform	DNA squencing system consisting of sequencing equipment (i.e. sequencer) and methodology to sequence target DNA. Examples include the Illumina MiSeq, the PacBio Single Molecular Real-Time (SMRT) and the Oxford Nanopore MinION sequencing platforms
Sequence read	A continuous DNA sequence determined from a target organism (e.g. bacterium)
Short sequence reads	A continuous DNA sequence determined from a target organism ranging between approximately 100-600 bp. For example, the Illumina MiSeq sequencing platform yields short sequence reads
Long sequence reads	A continuous DNA sequence determined from a target organism generally > 10 kb. For example, the PacBio SMRT sequencing platform yields long sequence reads
Sequence read error rate	The proportion of sequence reads containing sequencing errors
Contigs	Contiguous sequences assembled from overlapping smaller sequence reads that represent a consensus region of DNA
Single nucleotide variation (SNV)	A variation in a single nucleotide base that occurs at a specific position in the genome of an organism of interest without implying how often this variation occurs in a population
Single nucleotide polymorphism (SNP)	A variation in a single nucleotide base that occurs at a specific position in the genome of an organism of interest and is relatively common in a population
SNP analysis	Strain typing by mapping SNPs in sequence reads or assembled contigs against a reference genome(s). Reference genomes that are closely related to the sequenced samples are best
Core genome	A set of conserved genes present in virtually all isolates of a species.
Accessory genome	A variable set of genes present in some but not all isolates of a species. Examples in *S. aureus* and other staphylococci include SCC*mec* and arginine catabolic mobile elements (ACME)
Multilocus sequence typing (MLST). Also known as conventional MLST	A method that determines the sequences of internal fragments of up to seven housekeeping genes. The different sequences within a bacterial

Multilocus sequence typing (MLST). Also known as conventional MLST	A method that determines the sequences of internal fragments of up to seven housekeeping genes. The different sequences within a bacterial species are assigned as distinct alleles and, for each isolate, the alleles at each of the loci define the allelic profile or sequence type (ST). Curated MLST databases exist for many bacterial species including *S. aureus*. Conventional MLST has largely been replaced by WGS-based MLST schemes for *S. aureus* and other important bacterial pathogens due to significantly enhanced resolution
Whole genome MLST (wgMLST).	A typing method that uses WGS data from the core genome and accessory genome to perform MLST on a genome-wide basis and permits gene-by gene comparisons of very large numbers of genes amongst a group of isolates. The presence/absence of each target locus is determined from WGS data, as are allelic variants. Very high resolution can be achieved
Core genome MLST (cgMLST)	A typing method that uses WGS data to perform MLST on a predefined set of conserved genome-wide core (can be > 1000 genes) that are present in the vast majority of isolates of a bacterial species. Very high resolution can be achieved across large groups of isolates by core genome comparisons. The approach is highly reproducible across data sets
Genome assembly	The entire DNA sequence of an organism can be reassembled from overlapping sequence reads which first are assembled into larger contigs. Any remaining gaps between contigs can then be

Table II. Studies using whole genomic sequencing (WGS) of MRSA in the neonatal/paediatric setting

Country (Ref)	Study Details	Main Conclusions	Comment
Germany (35)	Retrospective case control study of neonates and staff	28 isolates formed a predominant strain with two other strains present Staff involved in transmission	Only one isolate per neonate/staff analysed and greater variability may have been present
Germany (36)	Outbreak of toxic-shock syndrome & enterotoxin-A producing MSSA	26% of neonates positive; 19.5% of staff 21 related isolates confirming PFGE analysis 2.8×10^{-6} mutations per nucleotide per year	Prolonged duration of outbreak suggests that staff member may have been reservoir as far back as 2-3 years ago
UK (37)	Retrospective investigation of outbreak in neonatal intensive care unit	Distinct cluster of outbreak isolates and clear separation with non-outbreak isolates.	Value of WGS in real time for MRSA control in hospitals
UK (38)	Outbreak in neonatal unit with isolates compared to other clinical strains in the hospital or community	11 MRSA infants identified ST2371 predominated Similar isolates in emergency department, out-patients & general practice	Mother-to-mother transmission outside hospital & staff carriage allowed outbreak to persist
USA (39)	Retrospective analysis of isolates from screening neonates, 2008-2010	Colonised infants had lower birth weight, gestational age and length of stay 70% of colonization due to events within the NICU. Transmission varied by strain	Useful longitudinal analysis of strains with evidence of significant within-unit spread and dynamic changes in strain pre-dominance
USA (40)	Temporal cluster of USA 300 identified on PFGE and analysed with WGS & *spa* typing	17 neonates acquired MRSA in unit 12 unique isolates with 5 in two clusters	WGS excluded cross-transmission in most cases, suggesting multiple and independent introduction of MRSA strains
Italy (41)	Isolates from putative outbreak in NICU & PICU	10/12 isolates ST625; 2xST8 A maximum 1.7% phylogenetic distance amongst ST625 A staff member isolated with ST625	Genetically related isolates but temporal analysis of isolates did not confirm an outbreak
	Isolates from two	Considerable homogeneity between 28	Highlights the importance of travel in the

Ireland (42)	outbreaks of MRSA ST78 in a neonatal intensive care unit	isolates with likely importation from abroad and involvement of healthcare worker	spread of MRSA and the value of WGS in tracking local spread and determining the origins of clones
France (43)	Isolates from two geographically separate NICUs	41 CC isolates from Limoges and Bordeaux with less than 22 SNPs difference	WGS is useful to determine relatedness but also to track bacterial evolution

MSSA; methicillin-susceptibility *Staphylococcus aureus*, PFGE; pulsed-field gel electrophoresis, NICU, neonatal intensive care unit, PICU; paediatric intensive care unit, CC, clonal complexes, SNPs, short nucleotide polymorphism.