American Journal of Infection Control 49 (2021) 59–64

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major Article

Molecular surveillance of methicillin-resistant *Staphylococcus aureus* genomes in hospital unexpectedly reveals discordance between temporal and genetic clustering

Rebecca Rose PhD [a,*], David J. Nolan BS [a], Samual Moot BS [a], Christopher Rodriguez [a], Sissy Cross BS [a], Yvette S. McCarter MD [b], Chad Neilsen MPH [c], Susanna L. Lamers MS [a]

[a] *BioInfoExperts, LLC, Thibodaux, LA*
[b] *Department of Pathology and Laboratory Medicine, UF Health Jacksonville, Jacksonville, FL*
[c] *Department of Infection Prevention & Control, UF Health Jacksonville, Jacksonville, FL*

Key Words:
Clustering
Molecular epidemiology
Phylogenetics
Hospital-acquired infections
Whole genome sequencing

Background: The objective of this study was to identify sources and linkages among methicillin-resistant *Staphylococcus aureus* infections using whole-genome sequencing (WGS).
Methods: A total of 56 samples were obtained from all patients with a confirmed MRSA infection over 6 months at University of Florida-Health Jacksonville. Samples were cultured and sequenced; data was analyzed on an automated cloud-based platform. Genetic Clusters were defined as <40 single nucleotide polymorphisms. Temporal Clusters were defined as ≥5 MRSA cases over 3 days.
Results: We found 7 Genetic Clusters comprising 15 samples. Four Genetic Clusters contained patients with non-overlapping stays (3-10 weeks apart), 3 of which contained patients who shared the same Unit. We also found 5 Temporal Clusters comprising 23 samples, although none of the samples were genetically related.
Discussion: Results showed that temporal clustering may be a poor indicator of genetic linkage. Shared epidemiological characteristics between patients in Genetic Clusters may point toward previously unidentified hospital sources. Repeated observation of related strains is also consistent with ongoing MRSA transmission within the surrounding high-risk community.
Conclusions: WGS is a valuable tool for hospital infection prevention and control.

BACKGROUND

Methicillin-resistant *Staphylococcus aureus* (MRSA) is one of the most important contributors to healthcare-associated infections (HAIs), resulting in ~20,000 deaths per year and thousands of dollars in extra costs.[1-3] Whole-genome sequencing (WGS) is increasingly being used by hospital infection control teams to elucidate infection dynamics (eg, whether cases are part of an internal outbreak of a single strain, have independent unrelated origins, or are associated with an ongoing community outbreak), which allows for a more objective determination of appropriate containment and treatment strategies. Furthermore, WGS in combination with traditional epidemiological analyses has both enabled accurate identification of previously unidentified sources of transmission in hospitals as well as ruled out the hospital as the source of the outbreak.[4-16] Importantly, the cost of WGS for a bacterial isolate is ~$300, making this an affordable and practical method for surveillance.

In this study, our objective was to use WGS to identify potential infection sources and outbreaks for 56 MRSA cases over a 6-month period at the University of Florida-Health Jacksonville (UFHJ). UFHJ is a large inner-city hospital network serving a high-risk urban population, and has a consistently elevated incidence of MRSA cases. As a comparison, we included WGS from a previous study which identified a MRSA outbreak in another Jacksonville hospital. In all, we analyzed a total of 158 samples using an automated, cloud-based analytical platform called FoxSeq (http://www.bioinfox.com), which is specifically designed to integrate the power of WGS analysis into hospitals.

---

[*] Address correspondence to Rebecca Rose, PhD, 718 Bayou Lane, Thibodaux, LA, 70301.

E-mail address: rebecca.rose@bioinfox.com (R. Rose).

Funding: RR, DJN, SM, CR, and SLL are supported by an NSF SBIR Phase II Award (PI SLL, #1830867).

Conflicts of interest: RR, DJN, SM, CR, SC, and SLL are full-time employees of Bioinfoexperts, LLC. RR, DJN, YSM, CN, and SLL conceived the study design. YSM and CN recruited patients. DJN performed the laboratory experiments. RR, DJN, SM, and CR developed the analytical platform. CN implemented the software in the IPCD. RR, DJN, SC, CN, and SLL interpreted results. RR, CN, and SLL wrote the manuscript.

https://doi.org/10.1016/j.ajic.2020.06.180

METHODS

Study site

UFHJ is a Level-1 trauma facility and part of the University of Florida Health System. UFHJ includes a 600-bed downtown UF Health Jacksonville hospital (“Downtown”), a newer 92-bed UFHJ North hospital located approximately 10 miles away (“North”), and over 20 Outpatient Clinics. The Downtown hospital is the major facility treating displaced, low-income and correctional populations, and serves as the “safety-net” for those that are turned away from other health care facilities. The North location serves communities that tend to be older and are more likely to be privately insured. UFHJ has historically seen high incidence rates of MRSA, despite intervention strategies including improved sepsis identification and management, more aggressive screening of at-risk populations (such as ICU admits), and improved sanitation protocols. The community served by UFHJ also typically has higher rates of MRSA at admission than in the general population, and many present with additional co-morbidities and/or engage in high-risk behaviors (eg, intravenous drug use [IVDU]).

Samples

The study was designed by the UFHJ Infection Prevention and Control Department (IPCD) to assess the use of WGS over a 6-month period (September 2018 - February 2019). During this time, 56 samples were collected from patients with active MRSA infections (as identified by a clinician and confirmed by the microbiology laboratory) from the 2 hospital campuses and ten clinics. Samples included blood, surgical site, urine, respiratory (sputum), tissue, and other fluids (eg, joints, abscesses, pericardial); nasal surveillance samples were not included. All patients had active infections that were treated with antibiotics. IRB approval was obtained under contract UF IRB201802582.

Sequencing

Frozen stocks of MRSA isolates were cultured overnight in 3 mL Luria broth at 37°C. 250 ml of culture was centrifuged at 1000 rpm for 10 minutes to pellet bacteria, and supernatant removed. Genomic DNA was extracted from bacterial pellets using the Qiagen DNeasy Blood and Tissue kit according to the manufacturer’s instructions. DNA concentrations were measured using an Invitrogen Qubit 2.0, and quality assessments were done on agarose gels. Sequencing libraries were prepared with the Illumina Nextera XT kit following the manufacturer’s recommendations for bead-based normalization. Up to 24 libraries were multiplexed on an Illumina miSeq V3 sequencing cartridge for a 2 × 300 bp paired-end run, with a 10% spike-in of Illumina PhiX control.

Typing and assembly

The sequence type (ST) for samples was determined computationally using *ariba,* which analyzes the raw reads to define ST based on the presence of specific mutations.[17] Raw read files were trimmed and filtered using *trimmomatic.*[18] Filtered reads were assembled to the appropriate NCBI reference genomes ST5 reference Mu50 (NCBI Reference Sequence: NC_002758.2) and ST8 reference USA300_FPR3757 (NCBI Reference Sequence: NC_007793.1) using *bowtie2*[19] with the default parameters. Mapped assemblies were formatted using *samtools*[20] and *picard* (http://broadinstitute.github.io/picard/).

Genetic variants

Genetic variants relative to the reference sequence were called using *HaplotypeCaller* and *GenotypeGVCFs* in GATK.[21] The core set of variants was defined as all sites with bi-allelic single nucleotide polymorphisms (SNPs) which had a quality score (QD)>20 and a maximum of 10% missing data across all samples. Filtered vcf files were processed using the *vcfR* and *apetools* packages in R (http://www.R-project.org) to obtain pairwise genetic distances. We also inferred maximum likelihood trees on the core set of SNPs using *RaxML*[22] with the GTRCAT model of nucleotide evolution and 100 bootstrap replicates to assess branch support. Scripts, output files (eg, vcf and distances) and outlier data (as described in Results below) are available at https://github.com/Bioinfoexperts/MRSA.

Genetic clustering

Samples were assigned to clusters using a “single linkage” method, where a sample was included in a cluster if the number of SNPs between it and any other sequence in the cluster is below the threshold.[23] Here we set a threshold of 40 SNPs, a commonly used threshold in molecular epidemiological studies of MRSA,[14,24-29] which is based on the observed intra-patient accumulation of mutations.[30] All analytics were performed on the FoxSeq automated pipeline on Amazon Web Services (AWS).

Comparison dataset

For additional resolution, 102 MRSA ST8 isolates from NCBI BioProject PRJNA341781 were included, which originated from 3 Jacksonville hospitals, primarily from their neonatal intensive care units (NICU): NICUA (n = 56, 2003 - 2011), NICUB (n = 17, 2010 - 2011), and NICUX (n = 29, 2010).[9] While these MRSA cases occurred several years prior (2003-2011), these isolates were sequenced at a similar depth using Illumina technology, and are from the same geographical area as the isolates included in this study.

RESULTS

Typing

Of the 56 UFHJ samples, *ariba* identified 42 samples as ST8, 5 samples as ST5, 4 samples as “novel,” and 5 samples as another unique subtype (the latter of which were excluded from further analysis). To maximize the number of samples included in the analysis, we grouped the 4 “novel” ST samples with the 42 ST8 samples. The UFHJ ST8 (n = 46) and ST5 (n = 5) datasets were analyzed separately.

Genetic distances

Genetic distances were calculated for every pair of ST8 and ST5 samples in the UFHJ dataset, and well as every pair of ST8 samples in the 3 comparison datasets (each hospital separately). Distances for each ST/hospital were plotted as boxplots. Several of the UFHJ ST8 samples were clearly outliers, in contrast to the other ST8 datasets. Upon further investigation, we determined that all distances >800 SNPs resulted from 4 ST8 UFHJ samples (2 of which typed as ST8 and 2 typed as “novel”). We excluded these 4 samples from the subsequent analyses. The remaining 42 UFHJ ST8 samples were then replotted (Fig 1). The interquartile range (IQR) for the UFHJ ST8 distances was higher than the IQR for the other 3 hospitals, although both NICUA and NICUX showed outliers (between 500 and 800 SNPs). The UFHJ ST5 IQR was much higher than any of the ST8 datasets, with a single outlier at the lower end of the distribution.

Temporal and genetic clusters

We used a 40 SNP threshold to define Genetic Clusters, which is a commonly used threshold in many genetic epidemiological studies of

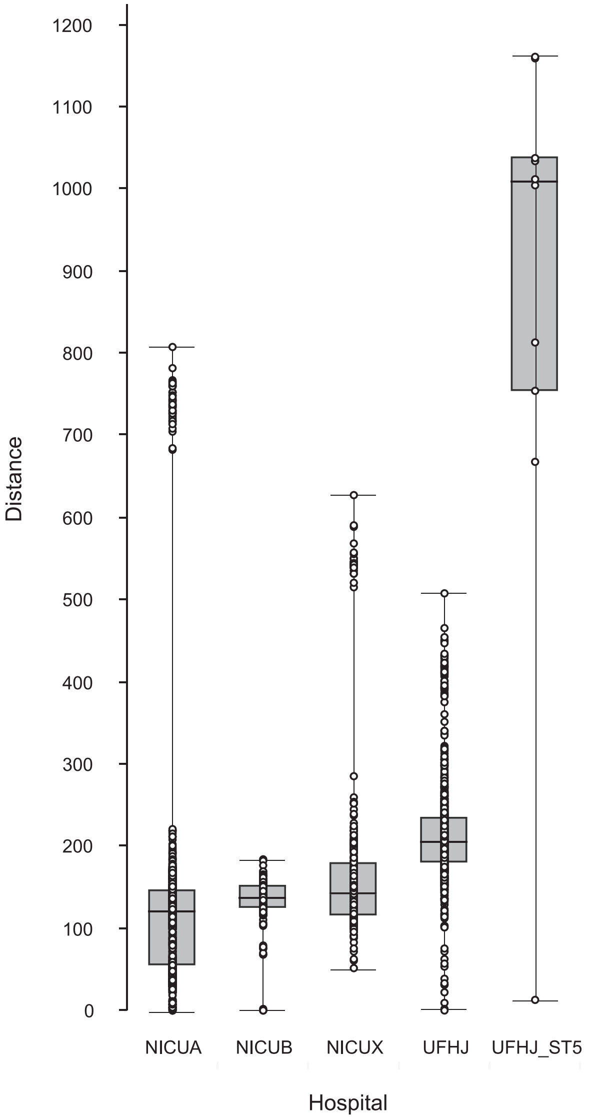

Fig 1. . Genetic distances among samples. The number of SNPs (y-axis) is shown for samples from each hospital (UFHJ is separated into ST8 and ST5). Samples are indicated by circles, and the filled light and dark grey boxes represent the upper and lower interquartile range (IQR). The whiskers extend the range of the data.

MRSA.[14,24-29] Genetic clustering was performed separately for the 42 UFHJ ST8 + Novel samples and the 5 UFHJ ST5 samples. We found a total of 7 Genetic Clusters (6 Genetic Clusters of ST8 samples, 1 Genetic Cluster of ST5 samples). Genetic Clusters were designated as #1 - #7, ordered according to the first observance at any of the 3 facilities.

The sampling and admission dates for all 56 samples were plotted, as well as the assignment of Genetic Clusters (Fig 2). Over the 6-month study period, 4 Temporal Clusters (1-4) were evident, defined as 5 or more cases of MRSA occurring within a 3-day time frame. At least 2 cases in each Temporal Cluster were at the Downtown site; however, none of the samples were genetically related to each other. On the other hand, in Temporal Clusters 2 and 4, 1 sample was genetically clustered with another sample which occurred earlier or later. Thus, WGS was a strong indicator that despite a high number of cases over a short time period, these were not true “outbreaks,” defined as multiple contemporaneous cases deriving from a single origin. On the other hand, 4 of the 7 Genetic Clusters (#1, #5, #6, and #7) comprised cases that were observed >3 weeks apart, suggesting that temporal clustering of cases is not a strong predictor for genetic relatedness.

Nonclustering cases

We then looked at the patients who did not cluster (temporally or genetically) with any other patient in the study (Fig 1). Interestingly, 5 patients showed symptoms at >2 days post-admission (DPA), which would be classified as HAIs. However, because none of these cases were genetically related to any other in the dataset, this suggests that these infections occurred in limited settings and did not circulate in the hospital. On the other hand, we previously noted several instances of genetically related infection re-appearing after an extended period of time, so we cannot rule out the possibility that the nonclustered cases are part of a Genetic Cluster occurring outside the 6-month window of this study.

Origin of genetic clusters

The UFHJ ICPD classifies infections as linked if the patients had overlapping stays in the same unit, and always classifies infections as HAIs if the patient developed symptoms >2 DPA. Patients who present with MRSA symptoms at the Outpatient Clinics are presumed to have a nonhospital acquired infection if they have not been previously hospitalized in the past 10 days. To determine how the genetic information informed the clinical definitions, we examined each Genetic Cluster in the context of additional epidemiological

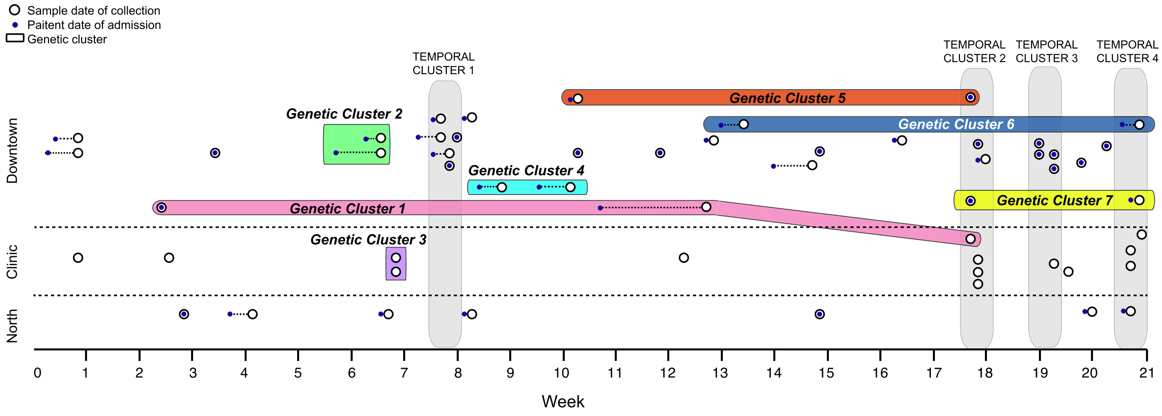

Fig 2. UFHJ MRSA cases plotted by sampling/admission dates. Each larger open dot represents a sample and the date of collection (typically equivalent to the date at which symptoms were first observed); each smaller closed dot indicates the date of admission; dotted lines connect the 2 dates for each sample. Samples are divided on the y-axis by the 3 locations. Samples classified into each of the 7 Genetic Clusters are grouped by colored bars. Four instances in which 5 or more cases were observed in a 3-day time frame are indicated by gray bars. (“Temporal Clusters”).

Table 1
Genetically clustered patient characteristics

Vial Number	Cluster	ST	Campus	Unit if applicable	Date of collection	Days post admission	Anatomic site	Gender	DOB Year	Risky community	Prior Hosp[*]
2800	1	8	Downtown	4N	2018/10/09	0	Blood	F	1985	IVDU	No
2881	1	8	Downtown	4N	2018/12/20	14	Wound	F	1997	IVDU	Yes[†]
2917	1	8	Clinic	NA	2019/01/24	NA	Wound	M	1983	IVDU	No
2823	2	5	Downtown	ICU	2018/11/07	6	Respiratory	F	1956	No	No
2828	2	5	Downtown	ICU	2018/11/07	2	Fluid, pericardial	F	1945	No	No
2827	3	8	Clinic	Breast	2018/11/09	NA	Fluid, breast	F	1960	No	No
2831	3	8	Clinic	Ear/Nose/Throat	2018/11/09	NA	Fluid, sinus	F	1956	No	No
2860	4	8	Downtown	P5S	2018/11/23	3	Wound	M	1981	Jail patient	No
2862	4	8	Downtown	7S	2018/12/02	4	Wound	M	1986	IVDU	No
2861	5	8	Downtown	ED	2018/12/03	1	Wound	F	1979	No	No
2906	5	8	Downtown	ED	2019/01/24	0	Blood	F	1980	No	No
2884	6	8	Downtown	6N	2018/12/25	3	Wound	M	1988	No	No
2968	6	8	Downtown	6N	2019/02/15	2	Wound	M	1980	No	No
2918	7	8	Downtown	ED	2019/01/24	0	Wound	M	1983	No	No
2970	7	8	Downtown	OR	2019/02/15	1	Wound	M	1982	IVDU	No

[*] Within past 10 days.
[†] At non-UFHJ facility.

information, including hospital/unit/bed locations, procedures, reason for admission, transfers, dates of admission and length of stay, additional risk factors, age, and gender.

Genetic Cluster 1 comprised 3 infections which spanned over 15 weeks, observed at 2 different facilities (Table 1). The patient initially presenting with this strain was already infected at admission and did not have prior hospitalization, and the patient presenting last was seen at the outpatient clinic with no prior hospitalization, both consistent with a community origin. On the other hand, MRSA symptoms for the second patient were observed at 14 DPA, suggesting an HAI. The first and second patient were on the same Unit (though nonoverlapping stays). Interestingly, all 3 patients were part of a high-risk community (IVDU). The 2 cases in Genetic Cluster 2 were detected on the same day, from the same Unit (ICU), and both were intubated and placed on a ventilator. In addition, these patients were infected with the less-common ST5 strain. All of these similarities are strong evidence for an HAI (Table 1). The 2 patients with Genetic Cluster 3 infections were seen in Outpatient Clinics, and because neither had prior hospitalizations in the previous 10 days, both were classified as CAIs. Both patients in the Genetic Cluster 4 had infections that were observed within 9 days of each other, and the patients shared clinical characteristics (location, anatomic site) and demographic characteristics (gender, similar age, and both from a high-risk group). Both infections were classified as HAIs based on DPA. Genetic Clusters 5 and 6 comprised 2 patients in the same facility and Unit, which might point towards an HAI. However, both sets of patients were sampled >7 weeks apart, and both patients in Genetic Cluster 5, and 1 in Cluster 6, showed symptoms <3 DPA (consistent with a CAI). Genetic Cluster 7 comprised 2 patients from the same facility (though not the same Unit) sampled 3 weeks apart, both of whom showed symptoms at 0 and 1 DPA (consistent with CAIs; Table 1).

Phylogenetics

To assess how the UFHJ ST8 infections related to those previously observed in the same city, an ML tree was inferred for the alignment of 23,317 SNPs for the 42 UFHJ ST8 samples and the 102 non-UFHJ ST8 samples. Because only 5 ST5 samples were observed in the UFHJ dataset, and none were observed in the comparison dataset, a phylogeny for the Genetic Cluster 2 samples was not inferred. Two non-UFHJ samples on long branches were removed to increase the visibility of the rest of the tree (Fig 3). In general, the samples from each of the 4 hospitals were interspersed in the tree. A major exception was a large clade containing 31 isolates, previously identified and which persisted for 2 years during the sampling duration.[9] The samples in the 6 UFHJ ST8 Genetic Clusters were evident on the tree with 100% bootstrap support. All 6 clusters showed very short branches within the clade, comparable to the branch lengths in the NICUA outbreak clade.

Several non-clustered UFHJ samples also grouped together on the tree with high bootstrap support, even though the samples were still separated by relatively long branches. However, the bootstrap support was generally high even for internal branches that grouped samples from multiple hospitals. Interestingly, an additional sample (2,921) grouped with the 2 Genetic Cluster 4 samples with 100% support and a relative short branch. Upon further investigation, the distance analysis showed that sample 2921 was separated from the other 2 samples by 54 and 58 SNPs, just beyond the cutoff for the Genetic Cluster definition.

DISCUSSION

In this study, we generated WGS from 56 MRSA cases that occurred over a 6-month time frame at locations in the UFHJ healthcare network. Seven clusters of genetically related UFHJ samples were identified, 6 of which comprised 2 samples, and 1 of which comprised 3 samples. Interestingly, the samples in only 2 of the 7 Genetic Clusters (#2 and #3) were observed within the same week; the samples in the other 5 Genetic Clusters were observed anywhere from 9 days to almost 7 weeks apart. In these latter instances, without the use of genetic information, these cases would almost certainly have not been classified as linked infections. Furthermore, there were 4 instances of Temporal Clusters (>4 MRSA cases within a 3-day time frame). In the absence of genetic evidence, these may have been considered an outbreak. However, the WGS analysis clearly showed that none of the cases in the Temporal Clusters were genetically related. Overall, these results showed that temporal clustering may be a poor indicator of genetic linkage.

We were also interested in whether using WGS could help define whether cases were of hospital or community origin, in addition to the timing of the observation of symptoms. For 1 Genetic Cluster (#2), the patients had a discordant classification based on DPA; however, there was strong evidence that both had an HAI (eg, same Unit and procedure). However, both patients presented with symptoms on the same day, so even in the absence of the genetic evidence, this overlap would still have triggered an investigation by the IPCD.

In another Genetic Cluster (#6), both patients were on the same Unit, although they had discordant classifications based on DPA and the infections were observed >7 weeks apart. While less straightforward, the shared Unit might still be a cause for concern both

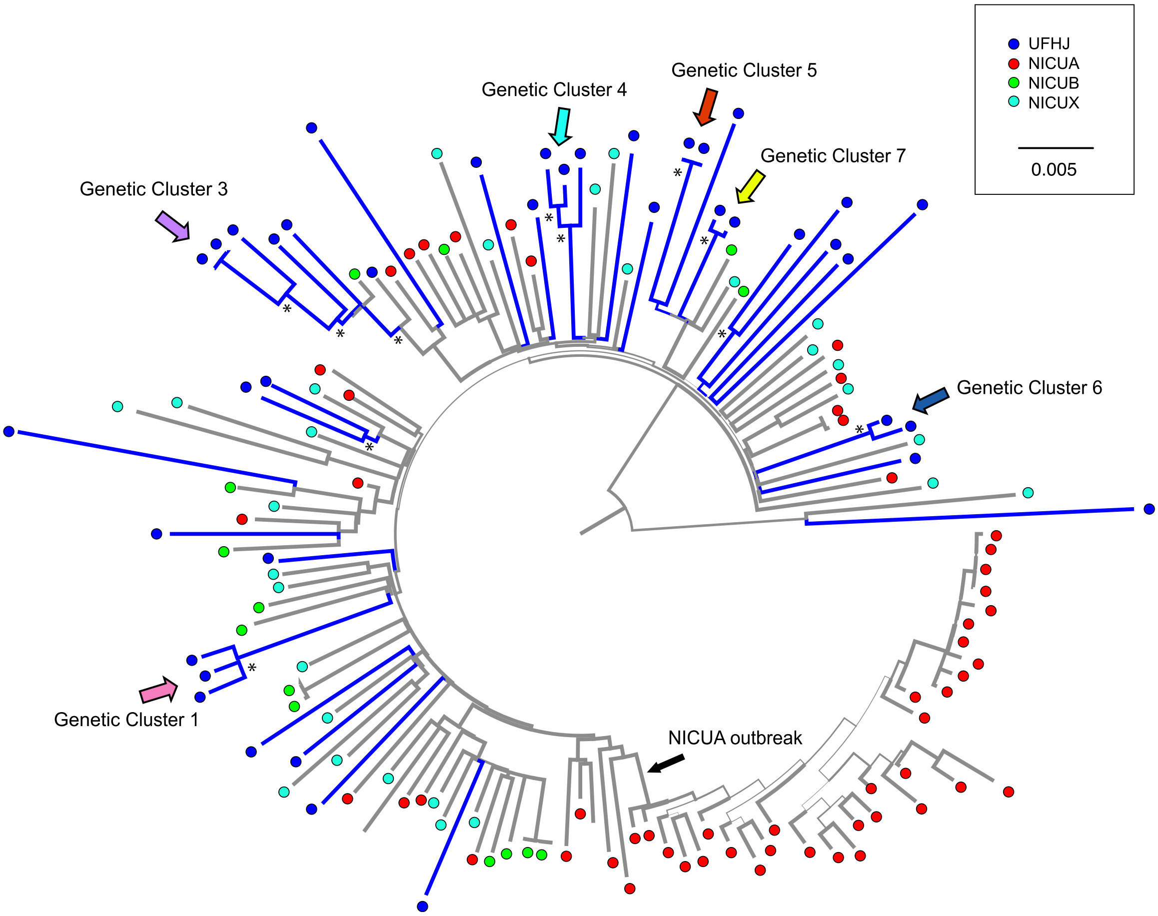

Fig 3. Maximum likelihood tree for ST8 samples (n = 142). Branches are proportional to genetic distance according to the legend. Branch thickness is proportional to the bootstrap support (thicker = higher) with (*) indicating branches of interest with 100% support. Dots at the tips represent samples; colors indicate the location of sampling according to the legend. Branches leading to UFHJ samples are colored blue. Arrows indicate UFHJ samples which clustered in the distance analysis.

infections were HAIs, and furthermore, that there was an undetected source of infection that persisted over a 2 month time frame. Similarly, both patients in Genetic Cluster #5 were classified as CAI based on DPA, yet both were in the same Unit (and again detected >7 weeks apart). While symptoms were detected sooner for both of these patients (0 and 1 DPA, respectively), this may also point toward an area of concern in the shared Unit.

On the other hand, for another Genetic Cluster (#1), the first and last patient were classified as CAIs and seen at different locations, while the second patient was classified as an HAI, showing symptoms 14 DPA. The second patient was on the same Unit as the first patient, thus providing a plausible epidemiological link for an HAI. On the other hand, all 3 patients were IVDU, which increases risk for a CAI, and the infections were observed over nearly a 3-month time frame. These factors taken together suggest a possibility that the second patient possibly acquired their infection from a community source, either prior to arrival in the hospital or from someone coming into the hospital, rather than from a nosocomial source. In addition, the observation that both Genetic Clusters #7 cases were CAIs, yet observed 3 weeks apart, suggests potentially sustained transmission of genetically-related strains of MRSA within the community, consistent with the high-risk urban population served by UFHJ.

Determining conclusively whether a particular infection was acquired in the hospital is obviously critical for both effective treatment/prevention, as well as for the financial implications of excessive HAIs for the hospital. WGS can provide valuable information about which strains are related; however, other lines of evidence are required to definitively determine the source of specific infections, eg, traditional epidemiological data and screening protocols. In the current study, determining the source (hospital vs community) for each Genetic Cluster is hampered by not knowing the MRSA status of each patient on admission. However, the results clearly point towards specific areas of concern, including a potential hospital source that may have been missed (Genetic Cluster #6), as well as a classified HAI with strong links to the community (Genetic Cluster #1). The most powerful use of WGS alone is to identify genetically unlinked cases, thereby excluding an outbreak scenario and/or a MRSA infection from a specific hospital-based source (eg, staff member or equipment).

Based on the results from this study, the UFHJ IPCD concluded that WGS provided valuable insights about the infection dynamics of MRSA; specifically, the critical observation that temporal clustering does not necessarily reflect true outbreaks (ie, multiple infections from the same source). Furthermore, the observation of genetically linked cases with shared epidemiological factors (eg, the same Unit), but spaced over 3+ weeks, suggests that a targeted intervention may be warranted in specific parts of the hospital. Ongoing monitoring of MRSA cases using WGS is expected to increase the resolution of linkage and enable the ICDs to make more informed decisions about prevention and care.

While WGS of pathogens has become an invaluable tool for public health departments,[31] it is not yet routine in hospitals. Limitations have included the significant computational resources and bioinformatics skills required for data analysis; however, innovations including cloud-based computing and integration with electronic health records are enabling hospital ICDs to incorporate this technology into surveillance protocols. In this study, all analyses here were automated and took place within protected AWS space, with results presented in a password-protected interactive web interface. This study has demonstrated that WGS and sophisticated bioinformatics tools such as FoxSeq can be implemented within ICPDs, and used by ICP team members without specialized bioinformatics training to better identify HAIs and improve patient care.

References

1. Kavanagh KT. Control of MSSA and MRSA in the United States: protocols, policies, risk adjustment and excuses. *Control Antimicrob Resist Infect Control*. 2019;8:103.
2. Kourtis AP, Hatfield K, Baggs J, Mu Y, See I, Epson E, et al. Vital signs: epidemiology and recent trends in methicillin-resistant and in methicillin-susceptible staphylococcus aureus bloodstream infections - United States. *MMWR Morb Mortal Wkly Rep*. 2019;68:214–219.
3. Klein EY, Jiang W, Mojica N, Tseng KK, McNeill R, Cosgrove SE, et al. National costs associated with methicillin-susceptible and methicillin-resistant staphylococcus aureus hospitalizations in the United States, 2010-2014. *Clin Infect Dis*. 2019; 68:22–28.
4. Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci Transl Med*. 2017;9.
5. Strauss L, Stegger M, Akpaka PE, Alabi A, Breurec S, Coombs G, et al. Origin, evolution, and global transmission of community-acquired Staphylococcus aureus ST8. *Proc Natl Acad Sci U S A*. 2017;114:E10596–E10604.
6. Popovich KJ, Snitkin ES, Hota B, Green SJ, Pirani A, Aroutcheva A, et al. Genomic and epidemiological evidence for community origins of hospital-onset methicillin-resistant staphylococcus aureus bloodstream infections. *J Infect Dis*. 2017; 215:1640–1647.
7. 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;55:2188–2197.
8. Tosas Auguet O, Stabler RA, Betley J, Preston MD, Dhaliwal M, Gaunt M, et al. Frequent undetected MRSA ward-based transmission linked to patient sharing between hospitals. *Clin Infect Dis*. 2017.
9. Azarian 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: e0164397.
10. 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: e0149844.
11. Roe CC, Horn KS, Driebe EM, Bowers J, Terriquez JA, Keim P, et al. Whole genome SNP typing to investigate methicillin-resistant Staphylococcus aureus carriage in a health-care provider as the source of multiple surgical site infections. *Hereditas*. 2016;153:11.
12. Harrison EM, Ludden C, Brodrick 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;8:102.
13. Azarian T, Cook RL, 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;36:777–785.
14. Price JR, Golubchik T, Cole K, Wilson DJ, Crook DW, Thwaites GE, et al. Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit. *Clin Infect Dis*. 2014;58:609–618.
15. Miller RM, Price JR, Batty EM, Didelot X, Wyllie D, Golubchik T, et al. Healthcare-associated outbreak of meticillin-resistant Staphylococcus aureus bacteraemia: role of a cryptic variant of an epidemic clone. *J Hosp Infect*. 2014;86:83–89.
16. Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis*. 2013;13:130–136.
17. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom*. 2017;3: e000131.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120.
19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
22. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–1313.
23. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJ, et al. Automated analysis of phylogenetic clusters. *BMC Bioinformatics*. 2013;14:317.
24. Long SW, Beres SB, Olsen RJ, Musser JM. Absence of patient-to-patient intrahospital transmission of Staphylococcus aureus as determined by whole-genome sequencing. *MBio*. 2014;5. e01692-14.
25. Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, Holden MTG, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat Commun*. 2015;6:6560.
26. Stine OC, Burrowes S, David S, Johnson JK, Roghmann MC. 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.
27. Popovich KJ, Snitkin E, Green SJ, Aroutcheva A, Hayden MK, Hota B, et al. Genomic epidemiology of USA300 methicillin-resistant Staphylococcus aureus in an urban community. *Clin Infect Dis*. 2016;62:37–44.
28. Chang HH, Dordel J, Donker T, Worby CJ, Feil EJ, Hanage WP, et al. Identifying the effect of patient sharing on between-hospital genetic differentiation of methicillin-resistant Staphylococcus aureus. *Genome Med*. 2016;8:18.
29. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, et al. Within-host evolution of Staphylococcus aureus during asymptomatic carriage. *PLoS One*. 2013;8:e61319.
30. Giulieri SG, Baines SL, Guerillot R, Seemann T, Gonçalves da Silva A, Schultz M, et al. Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent Staphylococcus aureus bacteraemia. *Genome Med*. 2018;10:65.
31. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS*. 2017;31:1211–1222.