ECCMID 2019 | Poster #P2686

Evaluation of the *Staphylococcus aureus* Analysis "1928D" Pipeline to Determine the Epidemiological Threshold Using Whole Genome Sequence Data

Lalitagauri M. Deshpande[1], Andrew P. Davis[1], F Fredrik Dyrkell[2], Dimitrios Arnellos[2], Rodrigo E. Mendes[1]
[1]JMI Laboratories, North Liberty, IA, USA; [2]1928 Diagnostics, Gothenburg, Sweden

Introduction

- *Staphylococcus aureus* is an important pathogen in the hospital and community settings, and the ability to correctly identify origins and genetic makeup of these organisms, including antimicrobial resistance, toxin portfolio and clonality, is essential for optimal treatment regimens
- Numerous tools are available for typing *S. aureus*, including pulsed-field gel electrophoresis (PFGE), multilocus sequencing typing (MLST), SCC*mec*-, *agr*- and *spa*-typing, and most recently, whole genome sequencing (WGS) analysis
- Traditional strain typing data always need to consider epidemiological concordance to determine relatedness while the sensitivity of WGS-based typing can help eliminate this necessity
- PFGE has long been the benchmark method; however, it has many challenges, including extensive time required to set up the tests and variation in the results obtained by different laboratories and/or technicians
- A pipeline platform that provides resistance mechanism predictions and core-genome (cg) epidemiological analysis for *S. aureus* has been developed
  - 1928 Diagnostics' *S. aureus* platform (1928D) uses an assembly-free method for identifying alleles, based on k-mer counting and k-mer coverage, in addition to targeted gene assembly performed to identify novel variants
- This study evaluated and compared the typing data provided by the 1928D platform against standard information utilized to differentiate persistent infection from re-infection among patients enrolled in a nosocomial pneumonia clinical trial

Materials and Methods

- Twenty pairs (n=40) of oxacillin (methicillin)-resistant *S. aureus* (MRSA) isolates (2 strains per patient) collected during baseline and follow-up visits from patients enrolled in a phase 4 global clinical trial were included
  - These isolates were characterized by applying standard PFGE, SCC*mec*-, *spa*- and *agr*-typing methods
- Isolates showing a PFGE profile (GelCompar II software, BioNumerics, Kortrijk, Belgium) similarity score of 100% were categorized as genetically identical, isolates with scores between >85% and <100% were considered genetically similar/related (subtype), and isolates exhibiting scores <85% were regarded as genetically distinct
- Additionally, 10 surveillance MRSA isolates possessing the same genetic background based on MLST data, but epidemiologically unrelated, were included as a control set
- All isolates were subjected to WGS on MiSeq (Illumina, San Diego, CA USA)
  - High-quality input DNA was extracted and purified using the KingFisher Cell and Tissue DNA kit (Thermo Scientific, Waltham, MA USA) in a robotic workstation KingFisher™ Flex Magnetic Particle Processor (Thermo Scientific)
  - DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and sequenced with a target depth of coverage >70X
  - Each raw data set was quality assured, error corrected, and assembled using SPAdes v. 3.9.0
- Raw reads (FASTQ files) were uploaded onto 1928 Diagnostics *S. aureus* platform (1928D) for analysis
  - SCC*mec* and MLST types as well as presence of genes encoding various toxins and resistance determinants were identified
  - Custom analysis was performed to determine allele differences and core genome single nucleotide polymorphisms (cgSNP) between isolate pairs
- The cgSNP and allelic variant results were compared to the PFGE reference method to determine thresholds for establishing identical, similar, or distinct relationships between baseline and follow-up isolates and also compared to the isolates from the control set based on MLST data
- Allelic distance and cgSNP analyses were also compared with MLST clonal complex (CC) results to add further granularity, especially between pairs of isolates from the same patient but showing different PFGE profiles

Results

- MLST, CC, and SCC*mec* (I, II, and IV) information obtained by standard methods matched data generated by 1928D in 28/30 clinical trial isolates, while most of the SCC*mec* III (8/10) isolates were unassigned (data not shown)
- Results comparing traditional typing methods to 1928D platform analyses of baseline and follow-up isolate pairs are summarized in Table 1 and Figure 1
  - Eleven patient sample pairs with respective identical epidemiological profiles showed the fewest (0 to 21 and 0 to 22) allelic and cgSNP differences
  - Four patient sample pairs with similar PFGE profiles (>85% and <100% similarity) showed slightly wider ranges (1 to 31 and 1 to 32) of allelic and cgSNP variations
  - Five isolate pairs with respective distinct genetic backgrounds demonstrated >200 differences
- Further analysis of genetically distinct isolate pairs (based on PFGE/conventional methods) showed much higher allelic and cgSNP distances (202 to 1268 and 219 to 9203, respectively)
  - An exception was found in patient "R", where both isolates belonged to CC5 (different PFGE and MLST), that showed only 202 and 219 differences respectively
- When the clinical isolate pairs of genetically related and epidemiologically unrelated controls were compared, allele and cgSNP distances of 45 to 244 and 46 to 277 were observed
  - These differences were higher than those observed within the epidemiologically related set (identical or similar PFGE profile) of clinical isolates (Table 1 and Figure 2)
  - CC5, a very diverse clonal type, exhibited the broadest range of allele and cgSNP differences (Figure 2)
- Sequence depth showed direct correlation with the fraction of core genes available for 1928D analysis (Figure 3)
- The 1928D analysis platform also provided antibiotic susceptibility predictions based on resistance genes/mutations detected; this data was not analysed as it was outside the scope of this study

Table 1 Summary of allele and cgSNP variations identified among isolates included in the study

	Analysis of clinical trial isolate pairs (no. of clinical isolates)			Analysis of epidemiologically unrelated CC-matched control set (no. of clinical and surveillance isolates)			
1928D analysis (difference in)	Identical (n=11)	Similar (n=4)	Different (n=5)	CC22 (n= 2 and 3)	CC5 (n= 11 and 4)	CC72 (n= 3 and 1)	CC8/239 (n= 6 and 2)
No. alleles	0-21	1-31	202-1268	57-65	45-244	114-119	66-230
No. cgSNP	0-22	1-32	219-9203	58-66	46-277	120-125	68-266

CC, clonal complex; cgSNP, core genome single nucleotide polymorphisms.

Table 2 Whole genome sequencing data analysis on 1928D platform of clinical isolate pairs categorized as epidemiologically unrelated by PFGE

Patient ID[a]	Isolate 1	Isolate 2	Allele distance	cgSNP distance	CC group	SCC*mec* types
E	10	32	1268	8128	5 and 72	II and IV
G	11	29	886	2493	8 and 5	III and II
N	1	20	1112	9203	5 and 8	II and III
R	14	15	202	219	5 and 5	IV and II
S	35	36	408	2548	8 and 5	IV and II

PFGE, pulsed-field gel electrophoresis; cgSNP, core genome single nucleotide polymorphisms; CC, clonal complex.

Figure 1 Pairwise analysis of patient samples classified by conventional methods and using whole genome sequencing data

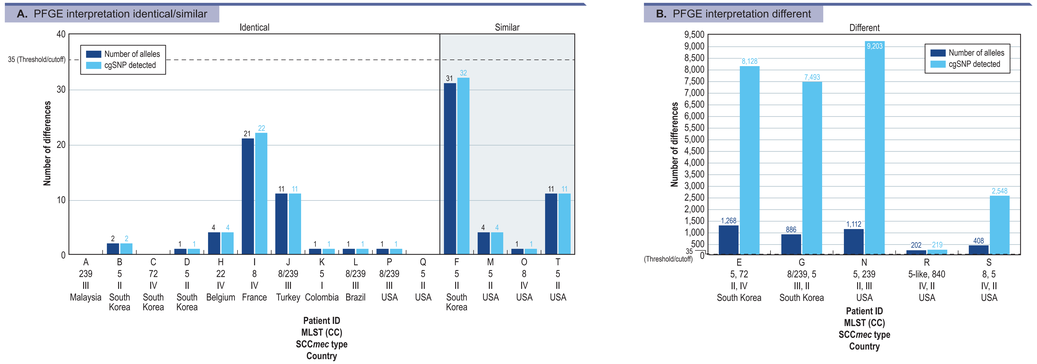
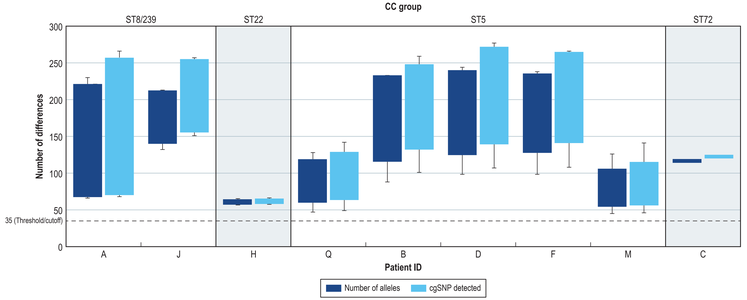
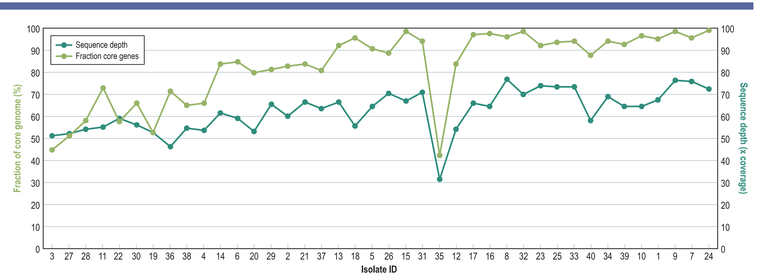


Figure 2 Pairwise analysis of genetically related and epidemiologically unrelated clinical samples



Boxplot represents range covered by 25th and 75th percentile, whiskers indicate minimum and maximum values of the data set

Figure 3 Comparison of whole genome sequencing depth with the fraction of core genome available for analysis on the 1928D platform



Conclusions

- The data suggest that, when combined with metadata, a threshold of approximately 35 (alleles and cgSNP differences) may be used for determining persistent versus re-infection cases in a clinical trial scenario
- 1928D is a robust platform for relatedness analysis of *S. aureus* that provides high-resolution typing information able to distinguish between isolates causing persistent infection (epidemiologically/genetically related) and re-infection cases (unrelated isolates)
- Although CC5 is a diverse group exhibiting the longest range of allele and cgSNP differences, genetically distinct isolates from the same patient belonging to this group showed the fewest differences in 1928D analysis
- Additional studies with a greater number of isolates are needed to refine this cutoff value

Acknowledgements

The authors would like to thank T.B. Doyle, V. Kantro and L. Flanigan at JMI Laboratories for assistance in the study and poster production.

References

Harris SR, Feil EJ, Holden MT, et al. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327: 469-474.

Mendes RE, Sader HS, Deshpande LM, et al. (2010). Characterization of baseline methicillin-resistant *Staphylococcus aureus* isolates recovered from Phase IV clinical trial for linezolid. *J Clin Microbiol* 48: 568-574.

Mendes RE, Deshpande LM, Smyth DS, et al. (2012). Characterization of methicillin-resistant *Staphylococcus aureus* strains recovered from a phase IV clinical trial for linezolid versus vancomycin for the treatment of nosocomial pneumonia. *J Clin Microbiol* 50: 3694-3702.

Monecke S, Coombs G, Shore AC, et al. (2011). A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. *PLoS One* 6: e17936.

Roe CC, Horn KS, Driebe EM, et al. (2016). 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* 153: 11.

Wang S, Wang Y, Shen J, et al. (2013). Polymorphic mutation frequencies in clinical isolates of *Staphylococcus aureus*: the role of weak mutators in the development of fluoroquinolone resistance. *FEMS Microbiol Lett* 341: 13-17.

Contact

Rodrigo E. Mendes, PhD
JMI Laboratories
345 Beaver Kreek Centre, Suite A
North Liberty, IA 52317
Phone: (319) 665-3370
Fax: (319) 665-3371
Email: rodrigo-mendes@jmilabs.com

To obtain a PDF of this poster:
Scan the QR code or visit https://www.jmilabs.com/data/posters/ECCMID19-1928Diagnostics-S.aureus-platform.pdf

Charges may apply. No personal information is stored.