bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Title:** A 7-member SNP Assay on the iPlex MassARRAY Platform Provides a Rapid and

2 Affordable Alternative to Typing Major African *Staphylococcus aureus* Types

3 **Running Title:** MassARRAY SNP genotyping of Kenyan clinical *S.aureus*

4 Justin Nyasinga,[b,d] Cecilia Kyany'a, [a] Raphael Okoth,[a] Valerie Oundo,[a] Daniel Matano,[a]

5 Simon Wacira,[a] Willie Sang,[c] Susan Musembi [b] and Lillian Musila[a#]

6 [a] United States Army Medical Research Directorate - Africa. P.O. Box 621-00606, Nairobi,

7 Kenya

8 [b] Kenyatta University. P.O. Box 43844-00100, Nairobi, Kenya

9 [c] Kenya Medical Research Institute, P. O. Box 54840-00200, Nairobi, Kenya

10 [d] Technical University of Kenya. P. O. Box 52428- 00200, Nairobi, Kenya

11 [#] Address correspondence to Lillian Musila, Lillian.Musila@usamru-k.org

12

13

14

15

16

17

18

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background: Data on the clonal distribution of *Staphylococcus aureus* in Africa is scanty, partly due to high costs and long turnaround times imposed by conventional genotyping methods such as *spa* and multilocus sequence typing (MLST) warranting the need for alternative typing approaches. This study applied and evaluated the accuracy, cost and time of using iPlex massARRAY genotyping method on Kenyan staphylococcal isolates.

Methods: Fifty four clinical *S. aureus* isolates from three counties were characterized using iPlex massARRAY, *spa* and MLST typing methods. Ten Single Nucleotide Polymorphisms (SNPs) from the *S. aureus* MLST database were assessed by iPlex massARRAY.

Results: The iPlex massARRAY assay grouped the isolates into 14 SNP genotypes with 9/10 SNPs interrogated showing high detection rates (average 89%). s*pa* and MLST typing revealed 22 *spa* types and 21 STs that displayed unique regional distribution. *spa* type t355 (ST152) was the dominant type and t2029 and t037 (ST 241) were observed among MRSA strains. MassARRAY showed 83% and 82% accuracy against *spa* and MLST typing respectively in isolate classification. Moreover, massARRAY identified all MRSA strains and a novel *spa* type. MassARRAY had reduced turnaround time (<12 hrs) compared to *spa* (3 days) and MLST (20 days) typing. The iPlex massARRAY cost approximately 18 USD compared to *spa* (30 USD) and MLST (126 USD) typing based on consumable costs/isolate.

Conclusion: Upon validation with a larger collection of isolates, iPlex massARRAY could provide a faster, more affordable and fairly accurate method of resolving African *S.aureus* isolates especially in large surveillance studies.

Key words: *S.aureus*, MRSA, typing, iplex massARRAY, *spa*, MLST, Kenya

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Antimicrobial resistance in *S. aureus* is a recognized global threat in clinical management of infections caused by the pathogen [1, 2]. The organism, which may exist asymptomatically in healthy individuals, is capable of causing a wide range of clinical syndromes including skin and soft tissue infections [3]. In Kenya, infections caused by *S. aureus* including methicillin resistant strains (MRSA) are widespread [4-6]. Recently, vancomycin resistant strains (VRSA) have been reported in Kenya's national referral hospital [7].

MRSA and VRSA strains are often associated with multi-drug resistance, the consequences of which include limited therapeutic options, long treatment periods, escalated treatment costs and significantly higher morbidity and mortality rates [2, 7]. Initially MRSA infections were associated with hospital settings such as surgical wards and the use of indwelling devices but community acquired strains (CA-MRSA) surfaced in the 1990s in Australia and North America and have now spread to all parts of the world [2, 8]. Other MRSA lineages (LA-MRSA) have been associated with livestock infections.

To curb drug resistance, adequate surveillance data regarding the emergence, dissemination and distribution of drug resistant strains is required. *spa* and MLST typing have become the most common molecular methods for characterizing the epidemiology of MRSA strains in surveillance studies [9, 10]. s*pa* typing is a sequence based technique that identifies the number and types of 24 nucleotide repeats in the variable X region of the staphylococcal protein A (*spa*) gene. The method is moderately discriminatory with interlaboratory repeatability and portability [10]. *spa* typing is faster and less expensive compared to MLST since it is a single-locus based technique. However the method presents challenges in delineating particular STs and cannot be

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reliably applied in outbreak detection [9]. MLST is a sequence based method that indexes polymorphisms in 400-500bp fragments of seven staphylococcal housekeeping genes. With moderate discriminatory power, interlaboratory portability and reproducibility the method is widely applied in surveillance [9]. However, the method is labour intensive, takes long and is costly with low-throughput given that seven loci are analyzed [9].

Particular strains of *S.aureus* tend to localize in certain geographical regions of the world for example the USA100, USA3OO in North America [2] and EMRSA15 and EMRSA16 in the United Kingdom [11]. In Kenya, only two hospital studies have reported on the molecular epidemiology of *S.aureus.* In one study, Omuse *et al*., reported the presence of major global MRSA clones such as MLST-CC5, 22, 7 and 30 among a heterogeneous collection of *S.aureus* isolates [6]. In the same study, members of MLST-CC5 such as ST1, ST5, ST8, ST241 were predominant. The other study conducted by Aiken *et al*., reported *spa* type t037-ST239 as the dominant strain among MRSA isolates recovered from inpatients in a mid-level hospital [5]. Generally, data on the clonal diversity and distribution of MRSA in Africa is scanty warranting further studies [1, 6, 12].

High costs, long result turnaround times, technical complexity and lack of technical expertise have impeded the adoption of conventional typing methods for surveillance studies especially in resource-limited settings such as Africa [1, 6, 12]. In recent years, SNP genotyping has been proposed as a robust, efficient and low cost approach to epidemiological characterization of *S.aureus* [9, 13]. Robertson *et al*., identified and interrogated by real-time PCR, seven informative SNPs that could resolve Australian MRSA strains into groups that reflected the population structure of the organism [14]. Syrmis *et al*., reported 98% accuracy and reduced assay costs when a multiplexed SNP-based assay on the iplex massARRAY platform was

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared to real-time PCR SNP genotyping of a collection of Australian *S.aureus* [13]. Despite the potential of iPlex massARRAY as a low cost, rapid and high throughput method for typing staphylococcal isolates, the technique has not been extensively evaluated against conventional methods like *spa* and MLST typing and against *S. aureus* clones from Africa whose epidemiology has been thought to be unique [12].

In this study, results of a multiplexed iPlex massARRAY assay for a collection of Kenyan *S.aureus* isolates are reported. In addition, comparisons between iPlex massARRAY, *spa* and MLST typing with regard to discriminatory power, assay turnaround times and reagent and consumable costs are presented.

MATERIALS AND METHODS

Study population and isolates. This study analyzed 54 archived clinical isolates collected between 2015 and 2016 from a recruited patient population of 221 persons. The isolates were part of an ongoing antimicrobial resistance surveillance project covering hospitals in three geographical regions of Kenya. The 54 isolates included 16 isolates from Nairobi County and 19 isolates each from Kisumu and Kericho Counties. Blood, urine, pus and swabs from skin and soft tissue infections and the throat were sampled from consenting in- and out- patients of all ages visiting the participating hospitals. Standard microbiological techniques were used for culture, isolation and identification of *S.aureus.* Isolate identities were further confirmed by conventional PCR for the *femA* gene. Antimicrobial susceptibility testing was performed using both Siemens MicroScan WalkAway 96 plus® (Siemens Healthcare Diagnostics, Inc., New York, USA) and Kirby Bauer disc diffusion methods. About 87% (47/54) of the isolates were associated with community acquired infections. Skin and soft tissue infections (SSTIs) accounted

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for 94% (51/54) of the infections while urine specimens represented 5.5% (3/54). Eleven percent (6/54) of the isolates were classified as MRSA. All MRSA isolates were *mecA*-positive and 4/6 had SCC*mec*IV genotype. All the isolates were preserved as glycerol stocks at -80^{0}C.

Ethical Approval. This study was reviewed and approved by the Kenya Medical Research Institute (KEMRI) Scientific and Ethics Review Unit (KEMRI/SERU/CCR/0061/3448) and by the WRAIR IRB (WRAIR #2089A).

DNA extraction. All clinical isolates and four reference isolates [ATCC 43300, ATCC CO-34, ATCC 25213 and ATCC 25923 (www.beiresources.org)] were sub-cultured by streaking a loopful of inoculum on Muller Hinton agar plates (Sigma-Aldrich, Missouri, USA) and incubating for 18-24hrs at 37^{0}C. The ZR/Fungal/ Bacterial DNA MiniPrep kit (Zymo Research, California, USA) was used for extraction following manufacturer's instructions with two modifications: The processing time for the cell disruption step was set at 10 mins and DNA was eluted with 60µl of elution buffer. DNA was quantified using Qubit DsDNA quantification kit (Thermo Fisher Scientific, Massachusetts, USA). For iPlex massARRAY assays, DNA concentrations were normalized to 10ng/µl. All DNA samples were stored at -20^0 C.

PCR amplification of the *spa* gene. Published primers were used to amplify the X region of the *spa* gene [15]. Each 25 µl PCR reaction mix included 10.5 µl of sterile nuclease free water, 12.5 µl of Dream Taq mix (Thermo Fisher Scientific, Massachusetts, USA), 0.5 µl (10pmol) of both the forward and reverse primers with 1 µl (approximately 50ng) of template DNA. Cycling was performed in a GeneAmp 9700 PCR System (Applied Biosystems, California, USA) with the following conditions: Initial denaturation at 95^{0}C for 5 mins, and 35 cycles of denaturation at 95^{0}C for 45s, primer annealing at 60^{0}C for 30s, extension at 72^{0}C for 90s followed by a final

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extension at 72^{0}C for 10 mins. Amplification products were resolved on a 1.5% agarose in 1X Tris Acetate EDTA buffer (Sigma-Aldrich, Missouri, USA) for 60 mins at 95V and visualized with EZ Vision DNA stain (Amresco Inc. Ohio, USA) using a UV- transilluminator (UVP LLC, California, USA).

PCR amplification of MLST loci. Published primers were used to amplify the MLST loci [16]. Each 25 µl reaction comprised 10.5 µl of sterile nuclease free water, 12.5 µl of Dream Taq mix (Thermo Fisher Scientific, Massachusetts, USA), 0.5 µl (10pmol) of both forward and reverse primers with 1 µl of template DNA (approximately 50ng). Thermocycler conditions were set as: Initial denaturation at 95^{0}C for 3 mins and 35 cycles of strand denaturation at 95^{0}C for 30s, primer annealing at 55^{0}C for 30s and extension at 72^{0}C for 60s and a final extension at 72^{0}C for 10 mins. The amplicons were resolved and visualized as described above for the *spa* gene.

Sanger sequencing of *spa* and MLST amplicons. *spa* and MLST amplicons were purified using the DNA Clean & Concentrator kit (Zymo Research, California, USA) following manufacturer instructions with a final elution volume of 30 µl. Both forward and reverse strands were sequenced. Each reaction contained 4 µl of sterile distilled water, 2 µl of 5X Big Dye buffer (Applied Biosystems, California, USA), 1 µl of the PCR primers (4 µM), 1 µl of Big Dye terminator mix (Applied Biosystems, California, USA) and 4 µl of amplicon DNA. Cycle sequencing was performed on an ABI 9700 thermocycler with cycling conditions set as: 94^{0}C for 5 mins followed by 30 cycles of 94^{0}C for 15s, 55^{0}C for 30s and 68^{0}C for 2 mins and 30s and a final extension of 68^{0}C for 3 mins. Sequencing fragments were purified using Sephadex G50 resin (Sigma-Aldrich, Missouri, USA) before loading on the Applied Biosystems Genetic Analyzer 3500.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Iplex massARRAY assays for MLST SNPs. The 10 MLST SNPs used in this study have been published before [13]. Nine of the 10 MLST SNPs assayed were bi-allelic while one (arcC210) was tri-allelic. Amplification and extension primers were designed using Agena Bioscience Assay Design Suite Version 2.0 (Agena Bioscience, Hamburg, Germany). The sequences used to design primers for primary amplification and extension PCR reactions were downloaded from the staphylococcal MLST database, http://www.mlst.net, 20th June, 2016 [16] and are shown in Table 1. All primary PCR primers had a 10-nucleotide tag (**ACGTTGGATG**) on the 5′ end to exclude them from the 4500-9000 Da mass range of MALDI-TOF MS detection. The primary multiplex PCR reactions were run on a GeneAmp 9700 PCR machine (Applied Biosystems, California, USA). Each reaction mix contained 1x PCR Buffer, 0.1 µM of forward and reverse primers, 4 mM $MgCl_2$, 500 µM dNTP mix, 0.5 units of Taq polymerase and 10 ng of DNA template. PCR conditions were set as follows: Initial denaturation at 95^0C for 2 minutes, 25 cycles of 95^0C for 30s, 56^0C for 30s, and 72^0C for 60s followed by a final extension at 72^0C for 5 mins.

Shrimp alkaline phosphatase (SAP) enzyme was used to dephosphorylate unused dNTPs from the primary PCR reaction before the second allele-specific primer extension PCR. For allele specific PCR reactions, 2 µl of the reconstituted extension primer cocktail was added to each reaction well and the following conditions were set: Initial denaturation at 94^0C for 30s, 40 cycles of one step at 94^0C for 5s with five sub cycles of 52^0C for 5s and 80^0C for 5s, and a final extension at 72^0C for 3 minutes. The sequences and masses for unextended primers (UEPs) and extension products for each SNP and SNP allele are shown in Table 2. Extension products were conditioned using a resin after which 10 nl was dispensed to a 96-well spectroCHIP using the MassARRAY Nanodispenser RS1000 (Agena Bioscience, Hamburg, Germany).

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Costing and time comparisons. Reagent and consumable costs per isolate for *spa* and MLST typing were estimated for various steps such as DNA extraction, primer sequences, PCR amplification, gel electrophoresis, PCR clean up, cycle sequencing, fragment purification and fragment analysis. For iPlex massARRAY, costs for primer sequences, primary and allele-specific PCR amplifications, SAP treatment, sample conditioning and liquid transfer to spectrochips were estimated. Time durations to result generation for the three methods after DNA extraction were also compared.

Data analysis. Raw sequencing chromatograms were examined using Chromas Version 2.6.2 (Technelysium Pty) before consensus sequences were created using BioEdit Sequence Alignment Editor Version 7.2.5 [17]. *spa* types were assigned using the online *spa* Type Finder/Identifier Software, *spa*typer.fortinbras.us/ (Fortinbras Research). *spa* types were further confirmed using the Ridom *spa* Server, *spa*server.ridom.de (Ridom GmbH, Würzburg, Germany) [18]. For MLST loci, consensus sequences were aligned using the online alignment tool, MAFFT Version 7 https://mafft.cbrc.jp/alignment/server/. Sequence types were assigned on the MLST database www.mlst.net [16] using the "Exact or Nearest Match" option. The SpectroAcquire program was used for data acquisition on the MassARRAY Compact Analyzer (Agena Bioscience, Hamburg, Germany) and detection parameters were set at ten laser shots per raster position with a threshold of five good spectra per sample pad.

Data availability. Raw chromatograms for *spa* and MLST typing as well as data output from the MALDI-TOF MS will be made available upon request.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESULTS

Distribution of *spa* and MLST types. *spa* and MLST typing of the collection of isolates yielded congruent results. *spa* and MLST typing showed 22 *spa* types and 21 STs respectively. The isolates displayed considerable heterogeneity with 16/22 *spa* types and 13/21 sequence types being represented by only one isolate. The three regions showed unique genetic fingerprints with minimal overlaps in isolate composition. Only t355 (ST152) was observed across the three regions with 18 *spa* types and 18 STs associating with specific regions. Kericho region showed greater diversity (12 *spa* types and 10 STs) followed by Kisumu (10 *spa* types and 8 STs) and Nairobi (5 *spa* types and 7 STs). Two novel *spa* types t17841 and t17826 were reported and were both associated with MSSA isolates. Despite the heterogeneity, t355 (ST152), t064 (ST8), t005 (ST22), t2029/t037 (ST241) seemed to dominate. The six MRSA isolates were represented by types: [t2029 (ST241) n=1, t037 (ST241) n=3, t355 (ST152) n=1 and t007 (ST39) n=1].

Typability and variability of the MLST SNPs. Table 3 summarizes the characteristics of the 10 SNPs analyzed. Nine of the 10 SNPs interrogated were highly typeable with SNP call rates (number of isolates in which a particular SNP was identified as a proportion of the total number of isolates tested) that ranged from 81% - 98.3% (average 89%). One SNP, pta294, was identified in only one isolate and was subsequently excluded from the analyses. Eight of the remaining SNPs were highly variable with allele frequencies (number of isolates positive for a given SNP allele as a proportion of all isolates in which that SNP was identified) ranging from 53% (arcC162) to 84.6% (yqil333).

Iplex massARRAY SNP genotypes. The combination of nine SNPs for each isolate was used to generate SNP genotypes. Different SNP combinations were created with the aim of finding the

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

219 smallest number of SNPs with the highest resolution (Data not shown). A 7-member SNP
220 classification that excluded both aroe252 and tpi36 **(arcC162, arcC210, aroE132, gmk129, tpi**
221 **241, tpi 243** and **yqil333)** achieved a similar resolution to that of the 9-member SNP profiles.
222 The assay grouped 44/54 isolates into 14 different SNP genotypes. Table 4 shows the regional
223 distribution of various SNP genotypes. In Kericho, nine SNP genotypes were observed among
224 the isolates while Kisumu and Nairobi isolates showed 8 and 4 SNP genotypes respectively. The
225 assay successfully typed 11/16, 16/19 and 17/19 isolates from Nairobi, Kericho and Kisumu
226 respectively while 10 isolates did not have complete SNP profiles. The classifications achieved
227 by iPlex massARRAY reflected the heterogeneity and regional distribution patterns revealed by
228 *spa* and MLST typing.

229 The 14 SNP genotypes were compared with corresponding *spa* and MLST sequence types for all
230 isolates. Iplex massARRAY showed 83% (34/41) accuracy against *spa* typing and 82% (32/39)
231 against MLST typing in genotype assignment. The SNP genotype **TTACGAC** corresponded to a
232 novel *spa* type t17841 and SNP genotype **TCACGGT** corresponded to another novel *spa* repeat
233 sequence yet to be assigned a *spa* type. iPlex massARRAY identified all MRSA isolates in this
234 collection. One MRSA was of the **ATACGGT** genotype which corresponded to t355. t037
235 which was the *spa* type of 3 MRSAs and t2029 (1 MRSA) were represented by the **ATACGAT**
236 genotype. **ACGTAGT** corresponded to t007 (one MRSA).

237 Three discrepancies involving iplex massARRAY were observed: SNP genotype **TCACGGC**
238 could not distinguish between t3772 and t13194 (ST 25 and ST80). SNP genotype **TTGTGGT**
239 could not distinguish between t314/ and t272 (ST121 and 152). Lastly, **TCACGAT** could not
240 distinguish between t084 and t131 (ST 15 and ST1290). Discrepancies involving *spa* types

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

t318/t021 and t037/t2029 were considered minor as such isolates are grouped as members of ST30 and ST241 respectively by MLST (Table 5).

Turnaround time, reagent and consumable costs per isolate. The iPlex massARRAY assay took approximately 1 day compared to *spa* and MLST typing which took approximately 3 and 20 days respectively (Table 6). The reagent and consumable cost of analyzing an isolate for massARRAY was approximately 18 USD compared to *spa* (30 USD/isolate) and MLST (126 USD/isolate) typing for the collection of isolates analyzed (Data not shown).

DISCUSSION

Seven MLST SNPs identified 14 unique genotypes and reflected the heterogeneity and distribution depicted by *spa*/MLST typing of a collection of 54 Kenyan isolates. iPlex massARRAY identified all the MRSA isolates used in this study as well as detecting a novel *spa* type. The assay showed comparable resolution to *spa* and MLST typing (83% and 82% respectively). The combination of 7 SNPs achieved similar resolution to that of 9 SNPs initially assayed. Coincidentally, the seven SNPs matched those originally described by Robertson *et al* that could resolve major Australian MRSA strains [14]. A technique that can identify circulating strains within a population using the smallest number of variable loci would offer the benefits of reduced costs and turnaround time. The isolate types identified by massARRAY such as ST152, ST5, ST8, ST15, ST30 and ST241 have been reported in other African countries such as Cameroon, Madagascar, Morocco, Niger, and Senegal [19] and Ghana [20] thereby highlighting the potential of this assay to be universally applied in the African context.

In some instances the assay could not distinguish between two or more *spa*/MLST types, an observation noted by others [14, 21]. The three SNP profiles that could not distinguish certain

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

263 *spa*/MLST types did not belong to major circulating *spa*/MLST types in the study as 5/6 affected
264 groups were represented by one isolate each, none of which was an MRSA strain. Previously, it
265 has been suggested that increasing the number of SNPs from 7 to 14 can resolve minor *S. aureus*
266 clones [21]. However, 2 of the 10 SNPs (aroE252 and tpi36) assayed did not prove useful in
267 increasing the resolution of the assay while one (pta294 was poorly typeable). While the
268 potential of a larger SNP set to resolve minor clones is recognized, a different set of SNPs in
269 addition to the seven identified here may be required.

270 MassARRAY demonstrated great capabilities for increased speed and throughput. After
271 optimization, the assay took <12 hours from DNA amplification to generate results for the
272 isolates analyzed in comparison to *spa* (3 days) and MLST typing (20 days). The technique
273 offers greater multiplexing capabilities that can upon effective assay design be used to detect
274 other genes to answer clinically important questions about a pathogen such as virulence and
275 resistance [13] a feature that is absent both in *spa* and MLST typing methods.

276 MassARRAY showed considerable reduction in assay costs in comparison to *spa* and MLST
277 typing. Estimated reagent and consumable cost/isolate was 18USD compared to *spa* and MLST
278 typing (30USD and 126 USD respectively). In one study, when the iplex massARRAY assay
279 costs were compared to SYBR green real time PCR, there was a 60% reduction in reagent costs
280 [13]. Trembizki *et al*. noted an approximately 30% reduction in cost compared to performing a
281 full MLST analysis [22]. The fairly high initial costs for the massARRAY system are justified by
282 the short turnaround times, multiplexing, automation and throughput capabilities which can
283 support multiple large scale studies concurrently thereby contributing to massive amounts of
284 data.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As the massARRAY technology is increasingly being adopted, studies utilizing its application in bacterial genotyping are being reported. Trembizki *et al* developed, applied and validated a 14-member SNP assay on the massARRAY platform for genetic characterization of *Neisseria gonorrhea* isolates in Australia [22]. Subsequently, the assay was used for a large scale AMR surveillance study of *Neisseria gonorrhea* [23].

The assay is promising particularly in long-term surveillance studies where it is impractical to perform *spa* and MLST typing on the hundreds to thousands of isolates recovered from such studies. An ideal approach would be to perform an iPlex massARRAY analysis as a first-step molecular screen to identify isolates with unusual SNP genotypes or those with which discrepant results are recognized. A full *spa* or ST determination could then follow to definitively resolve such isolates.

There were instances for example with pta294 where particular SNPs could not be called. A possible explanation for this could be sequence variation in the primer targets for the primary PCR reactions, an observation that has been noted elsewhere [13, 22]. The quality of DNA can potentially affect the success of the assay [22]. However, the DNA used for these experiments was of high quality as extraction was done using a commercial kit and DNA concentrations were measured and normalized.

The lack of a database for MLST SNPs that can be used for matching particular SNP profiles to known STs or *spa* types [22] is a major limitation. This is partly due to the fact that assays on this platform have not been rigorously evaluated against known STs/*spa* types from an international collection of isolates. However, with increased validation and adoption, it should be

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

306 possible to have an online MLST-style platform where SNPs can be submitted for inter-
307 laboratory comparability of data.

308 Compelling epidemiological conclusions cannot be drawn as the study analyzed a modest
309 collection of 54 isolates, a small proportion of which constituted MRSA isolates. A larger
310 collection of isolates from diverse regions and clinical syndromes would give not only a reliable
311 reflection of the epidemiology of MRSA in Kenya but also serve to highlight the utility of iplex
312 massARRAY for surveillance.

313 In conclusion, seven SNPs derived from the MLST loci provided comparable discriminatory
314 power for resolving a heterogeneous and regionally unique collection of Kenyan clinical
315 *S.aureus* isolates including MRSA strains. The iplex massARRAY demonstrated advantages of
316 reduced turnaround time and assay costs in relation to two conventional typing methods. With
317 increased validation, the assay should serve as a complement to existing typing methods
318 especially in staphylococcal AMR surveillance studies.

319 **Conflict of Interest:** The authors declare no conflict of interest.

320 **Acknowledgements:** This work was funded by the Armed Forces Health Surveillance Branch
321 (AFHSB) and its Global Emerging Infections Surveillance (GEIS) Section. PROMIS ID:
322 20160270153 FY17. JN received a student research grant from the Kenya National Research
323 Fund. We acknowledge the use of the *S. aureus* MLST database which is located at Imperial
324 College London and is funded by the Wellcome Trust. This publication made use of the *spa*
325 typing website (http://www.spaserver.ridom.de/) that is developed by Ridom GmbH and curated
326 by SeqNet.org (http://www.SeqNet.org/). We appreciate technical support from Luiser Ingasia

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and the MDR Laboratory. We thank the study participants for the AMR Surveillance Project. This article has been published with permission from the Director, KEMRI.

Disclaimer: Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70-25.

Author contributions: JN: Study design, strain typing, data analysis and manuscript write up and review; **LM:** Study design, data analysis, manuscript write up, review and proof reading; **CK:** DNA extraction, strain typing and manuscript review; **RO:** MassARRAY assay design and execution, data analysis and manuscript review. **VO, DM** and **SW** performed bacterial culture and identification, AST testing and manuscript review; **WS** and **SM** offered technical consultation and manuscript review.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Abdulgader SM, Shittu AO, Nicol MP, Kaba M (2015). *Molecular epidemiology of Methicillin-resistant Staphylococcus aureus in Africa: a systematic review.* Front Microbiol **6**: p. 348.

2. DeLeo FR, Otto M, Kreiswirth BN, Chambers HF (2010). *Community-associated meticillin-resistant Staphylococcus aureus.* Lancet **375**(9725): p. 1557-1568.

3. Loughman JA, Fritz SA, Storch GA, Hunstad DA (2009). *Virulence gene expression in human community-acquired Staphylococcus aureus infection.* J Infect Dis **199**(3): p. 294-301.

4. Akoru C, Kuremu RT, Ndege SK, Obala A, Smith JW, Bartlett M (2016). *Prevalence and Anti-Microbial Susceptibility of Methicillin Resistant Staphylococcus aureus at Moi Teaching and Referral Hospital Eldoret.* Open J Med Microbiol **6**(01): p. 9.

5. Aiken AM, Mutuku IM, Sabat AJ, Akkerboom V, Mwangi J, Scott JA, Morpeth SC, Friedrich AW, Grundmann H (2014). *Carriage of Staphylococcus aureus in Thika Level 5 Hospital, Kenya: a cross-sectional study.* Antimicrob Resist Infect Control **3**(1): p. 22.

6. Omuse G, Van Zyl KN, Hoek K, Abdulgader S, Kariuki S, Whitelaw A, Revathi G (2016). *Molecular characterization of Staphylococcus aureus isolates from various healthcare institutions in Nairobi, Kenya: a cross sectional study.* Ann Clin Microbiol Antimicrob **15**(1): p. 51.

7. Gitau W, Masika M, Musyoki M, Museve B, Mutwiri T (2018). *Antimicrobial susceptibility pattern of Staphylococcus aureus isolates from clinical specimens at Kenyatta National Hospital.* BMC Res Notes **11**(1): p. 226.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8. Chambers HF, DeLeo FR. (2009). *Waves of resistance: Staphylococcus aureus in the antibiotic era.* Nat Rev Microbio **7**(9): p. 629.

9. Struelens M, Hawkey PM, French G, Witte W, Tacconelli E (2009). *Laboratory tools and strategies for methicillin-resistant Staphylococcus aureus screening, surveillance and typing: state of the art and unmet needs.* Clin Microbiol Infect **15**(2): p. 112-119.

10. O'Hara FP, Suaya JA, Ray GT, Baxter R, Brown ML, Mera RM, Close NM, Thomas E, Amrine-Madsen H (2016). *spa Typing and multilocus sequence typing show comparable performance in a macroepidemiologic study of Staphylococcus aureus in the United States.* Microb Drug Resist. **22**(1): p. 88-96.

11. Holmes A, McAllister G, McAdam P, Choi SH, Girvan K, Robb A, Edwards G, Templeton K, Fitzgerald J (2014). *Genome-wide single nucleotide polymorphism-based assay for high-resolution epidemiological analysis of the methicillin-resistant Staphylococcus aureus hospital clone EMRSA-15.* Clin Microbiol Infect **20**(2): p. O124-O131.

12. Schaumburg F, Alabi A, Peters G, Becker K (2014). *New epidemiology of Staphylococcus aureus infection in Africa.* Clin Microbiol Infect **20**(7): p. 589-596.

13. Syrmis MW, Moser RJ, Whiley DM, Vaska V, Coombs GW, Nissen MD, Sloots TP, Nimmo GR (2011). *Comparison of a multiplexed MassARRAY system with real-time allele-specific PCR technology for genotyping of methicillin-resistant Staphylococcus aureus.* Clin Microbiol Infect **17**(12): p. 1804-1810.

14. Robertson GA, Thiruvenkataswamy V, Shilling H, Price EP, Huygens F, Henskens FA, Giffard PM (2004). *Identification and interrogation of highly informative single*

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. J Med Microbiol **53**(1): p. 35-45.

15. Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, Bost DA, Riehman M, Naidich S, Kreiswirth BN (1999). *Evaluation of protein A gene polymorphic region DNA sequencing for typing of Staphylococcus aureus strains.* J Clin microbiol **37**(11): p. 3556-3563.

16. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000). *Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus.* J Clin microbiol **38**(3): p. 1008-1015.

17. Hall TA (1999). *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT* in *Nucleic acids symposium series*. [London]: Information Retrieval Ltd., c1979-c2000.

18. Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, Vogel U (2003). *Typing of methicillin-resistant Staphylococcus aureus in a university hospital setting by using novel software for spa repeat determination and database management.* J Clin microbiol **41**(12): p. 5442-5448.

19. Breurec S, Fall C, Pouillot R, Boisier P, Brisse S, Diene-Sarr F, Djibo S, Etienne J, Fonkoua MC, Perrier-Gros-Claude, JD, Ramarokoto CE, Randrianirina F, Thiberge JM, Zriouil SB, Working Group on *Staphylococcus aureus* Infections, Garin B, Laurent F (2011). *Epidemiology of methicillin-susceptible Staphylococcus aureus lineages in five major African towns: high prevalence of Panton-Valentine leukocidin genes.* Clin Microbiol Infect **17**(4): p. 633-639.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Egyir B, Guardabassi L, Sørum M, Nielsen SS, Kolekang A, Frimpong E, Addo KK, Newman MJ, Larsen AR (2014). *Molecular epidemiology and antimicrobial susceptibility of clinical Staphylococcus aureus from healthcare institutions in Ghana.* PLoS One **9**(2): p. e89716.

21. Stephens AJ, Huygens F, Inman-Bamber J, Price EP, Nimmo GR, Schooneveldt J, Munckhof W, Giffard PM (2006). *Methicillin-resistant Staphylococcus aureus genotyping using a small set of polymorphisms.* J Med Microbiol **55**(1): p. 43-51.

22. Trembizki E, Smith H, Lahra MM, Chen M, Donovan B, Fairley CK, Guy R, Kaldor J, Regan D, Ward J, Nissen MD, Sloots TP, Whiley DM (2014). *High-throughput informative single nucleotide polymorphism-based typing of Neisseria gonorrhoeae using the Sequenom MassARRAY iPLEX platform.* J Antimicrob Chemother, **69**(6): p. 1526-1532.

23. Whiley DM, Trembizki E, Buckley C, Freeman K, Baird RW, Beaman M, Chen M, Donovan B, Kundu RL, Fairley CK, Guy R, Hogan T, Kaldor JM, Karimi M, Limnios A, Regan DG, Ryder N, Su JY, Ward J, Lahra MM (2017). *Molecular antimicrobial resistance surveillance for Neisseria gonorrhoeae, Northern Territory, Australia.* Emerg Infect Dis **23**(9): p. 1478.

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

435 **Table 1:** Primary amplification primer sequences for various SNPs and their expected amplicon

436 sizes

SNP ID*	Forward primer sequence (5′-3′)**	Reverse primer sequence (5′-3′)**	Amplicon length (bp)
arcC162	GTTGGGTTATCAAATCGTGG	ATAGTGATAGAACTGTAGGC	96
arcC210	GAGTCTGGCTGTTCTTTTTG	GATAAAGATGATCCACGATTC	118
aroE132	GGCTTTAATATCACAATTCC	GCACCTGCATTAATCGCTTG	103
aroE252	TCTGGATAAACGCTGTGCAA	AAGATGGCAAGTGGATAGGG	93
gmk129	AACTAGGGATGCGTTTGAAG	GGTGTACCATAATAGTTGCC	104
pta294	GATTAGTAAGTGGTGCAGCG	CCAGATGTTTTTGTAACGCC	114
tpi36	ATTCCACGAAACAGATGAAG	ACGCTCTTCGTCTGTTTCAC	120
tpi241	ATGAGCCAATCTGGGCAATC	GTTTGACGTACAAATGCACAC	102
tpi243	TGAACCAATCTGGGCAATCG	GTTTGACGTACAAATGCACAC	101
yqil333	CAACCATTGATTGATGTCCC	GGCGGTATGGAGAATATGTC	105

437

438 ***arcC:*** Carbamate kinase, **aroE:** Shikimate dehydrogenase, **gmk:** Guanylate kinase, **pta:**

439 Phosphate acetyl transferase, **tpi:** Triosephosphate isomerase, **yqil:** Acetyl co-enzyme A acetyl

440 transferase. ** All amplification primers had a 5′ 10-mer tag.

441 **Table 2:** Allele-specific extension primer sequences and expected masses for various SNP

442 alleles

SNP ID	Alleles	Unextended primer (UEP) sequence (5′-3′)*	UEP mass	Call 1	Mass 1 (Da)	Call 2	Mass 2 (Da)
arcC162	T/A	CTGTAGGCACAATCGT	4881.2	A	5152.4	T	5208.3
arcC210	C/T/A	cTCCACGATTCAATAACCCAAC	6592.3	C	6839.5	T	6919.4
aroE132	A/G	agGATCTAAATACGGTATGATACG	7424.9	G	7672	A	7752
aroE252	T/A	ACAGATGGTATTGGTTATGT	6202	A	6473.3	T	6529.1
gmk129	C/T	CAGCATATTCTATAAATTGGTCATCTTT	8527.6	T	8798.8	C	8814.8
pta294	A/C	TGCTGGACGTACAGTATC	5514.6	C	5801.8	A	5841.7
tpi36	C/T	agCATTCCATGTTTGAAAATAGC	7046.6	T	7317.8	C	7333.8
tpi241	G/A	AATGCACACATTTCATTTG	5761.8	G	6009	A	6088.9
tpi243	A/G	CAAATGCACACATTTCATT	5730.8	G	5978	A	6057.9
yqil333	C/T	gctgGTCAACAACAGTCGCTT	6406.2	C	6653.4	T	6733.3

443

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

444 * The bases in lowercase are incorporated into the oligonucleotide sequences during primer
445 design to prevent the extension products for various SNP alleles from being too close to each
446 other in the mass spectrum.

447 **Table 3:** Summary of the 10 MLST SNPs analyzed by iPlex massARRAY

SNP	Polymorphism	SNP call rate [a]	Allele frequency [b]
arcC162	T/A	98.3% (57/58)	**T** 53% (30/57)
arcC210	C/T/A	86% (50/58)	**T** 60% (30/50)
aroE132	A/G	91.3% (53/58)	**A** 73.6% (39/53)
aroE252	T/A	86% (50/58)	**T** 64% (32/50)
gmk129	C/T	84.5% (49/58)	**C** 67.3% (33/49)
tpi36	C/T	81% (47/58)	**C** 63.8% (30/47)
tpi241	G/A	86% (50/58)	**G** 98% (49/50)
tpi243	A/G	94.8% (55/58)	**G** 60% (33/55)
yqil333	C/T	89.7% (52/58)	**T** 84.6% (44/52)
pta294	A/C	- *	-*

448

449 * pta294 was identified in only one isolate and was excluded when generating SNP genotypes.
450 [a] **SNP call rate**: The proportion of isolates in which a given SNP was identified as a proportion
451 of all isolates tested for that SNP. [b] **Allele frequency:** The number of isolates in which a given
452 SNP allele was identified as a proportion of all isolates in which the SNP was identified.

453

454

455

456

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

457 **Table 4:** Regional distribution of *S. aureus* isolates by massARRAY SNP genotyping

SNP genotype*	Nairobi	Kericho	Kisumu	Total
ATACGGT	6	5	1	12
TCGCGGT	3	1		4
TCACGAT	1	2		3
TTACGAT	1		6	7
TCACGGT		1		1
TTACGGT		1		1
ATGTGGT		3		3
TTATGAC		1		1
TTACGAC		1	1	2
TTGTGGT		1	1	2
ATACGAT			4	4
ACGTAGT			1	1
TCATGAC			1	1
TCACGGC			2	2
Incomplete profiles**	5	3	2	10
Total	16	19	19	54

458

459 ***7- member SNP genotypes are in the order of **arcC162, arcC 210, aroE 132, gmk 129, tpi**

460 **241, tpi 243** and **yqil 333**. ****:** Isolates in which at least one of the seven SNPs used for genotype

461 assignment was not identified were considered incomplete.

462 **Table 5:** A summary of comparisons between iplex massARRAY, *spa* and MLST typing

SNP genotype	*spa* type	Sequence type
ATACGGT	t355	ST152
TCGCGGT	t005	ST22
TTACGAT	t064	ST8
TTACGAC	t17841*	ST121
TTATGAC	t189	ST188
ACGTAGT	t007	ST39

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TCACGGT	**	ST1925
TCATGAC	t002	ST5
ATGTGGT	t318/t021	ST30
ATACGAT	t037/t2029	ST241
TTACGGT	t10499	ST 2019
TCACGGC	t3772/ t13194	ST 25/ST80
TTGTGGT	t314/ t272	ST121/152
TCACGAT	t084/ t131	ST 15/ST1290

463

464 7-member SNP genotypes are in the order of **arcC162, arcC 210, aroE 132, gmk 129, tpi 241,**

465 **tpi 243** and **yqil 333**. * A novel *spa* type reported in this study ** An unassigned novel *spa*

466 repeat sequence

467 **Table 6:** Time-to-result analysis for *spa*, MLST and massARRAY typing methods for the

468 isolates after DNA extraction

	***spa* typing**		**MLST typing**	
Step	**Time**	**Period**	**Time**	**Period**
Amplification PCR	3 hrs	1 day	21 hrs	2 days
Gel electrophoresis	2 hrs		7 hrs	1 day
PCR amplicon clean up	2 hrs		14 hrs	2 days
Sequencing PCR	3 hrs	1 day	21 hrs	4 days
Pre-sequencing purification	3 hrs		21 hrs	8 days
Sequence fragment analysis	10 hrs		70 hrs	
Data analysis	7 hrs	1 day	30 hrs	3 days
Total time		3 days		20 days
iPlex massARRAY				
Step	**Time**	**Period**		
Amplification PCR	2.5hrs	1 day		
SAP treatment	0.75hrs			
iPlex PCR	3 hrs			
Sample conditioning	1 hr			

bioRxiv preprint doi: https://doi.org/10.1101/395079; this version posted August 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data acquisition	1 hr		
Data analysis	3 hrs		
Total time		1 day	

469

470

471