

1 **Full Title:** Strain typing and characterization of virulence genes in clinical *Staphylococcus aureus*  
2 isolates from Kenya.

3 Short title: *S.aureus* strain typing and virulence gene detection.

4 **Authors**

5 Cecilia Kyany'a<sup>1, 2</sup>, Justin Nyasinga<sup>3</sup>, Daniel Matano<sup>1, 2</sup>, Valerie Oundo<sup>1, 2</sup>, Simon Wacira<sup>1, 2</sup>, Willie  
6 Sang<sup>1</sup> and Lillian Musila<sup>1, 2, \*</sup>

7 **Author affiliations**

8 <sup>1</sup> Kenya Medical Research Institute, P. O. Box 54840 - 00200, Nairobi, Kenya

9 <sup>2</sup> The United States Army Medical Research Directorate-Kenya, P.O. Box 606-00621, Nairobi,  
10 Kenya

11 <sup>3</sup> Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya

12 \*Corresponding author

13 Email: [Lillian.musila@usamru-k.org](mailto:Lillian.musila@usamru-k.org)

## 14 **Abstract**

15 *Staphylococcus aureus* strain typing is an important surveillance tool as particular strains have  
16 been associated with virulence and community and hospital acquired MRSA outbreaks globally.  
17 This study sought to determine the circulating strain types of *S.aureus* in Kenya and establish  
18 the virulence genes among the strains. Clinical *S.aureus* isolates from 3 hospitals in Kenya were  
19 sequenced on the Illumina Miseq and genomes assembled and annotated on PATRIC. Results  
20 demonstrated great diversity among the isolates with identification of 6 distinct CC  
21 (8,22,15,80,121,152), 8 ST types (8, 15, 22,80,121,152,241, 1633) and 8 spa types (t005, t037, t064,  
22 t084, t233, t2029, t272,t355). Novel STs (4705, 4707) and a novel spa type (t17826) were  
23 identified. The most prominent clonal complex was CC 152 comprised of only MSSA. A majority  
24 of MRSA isolates (3/4) typed to ST 241, CC8. One MRSA isolate typed to a novel ST 4705. All  
25 isolates were screened for a panel of 56 known virulence genes (19 adhesins, 9 hemolysins, 5  
26 immune evasion proteins, 6 exo-enzymes and 19 toxins). 9 toxin genes were detected among  
27 the isolates with CC8 isolates having the highest numbers of toxin genes. An MSSA isolate (CC8)  
28 from a severe burn infection had the highest number of toxin genes (5). All MRSA isolates (CC8)  
29 had only 2 toxins, SEK and SEQ, whereas a majority of the MSSA isolates either had 0 or  $\geq 2$   
30 toxins. SEK+SEQ and TSST-1+SEB+SEL toxin combinations were observed among patients  
31 whose disease resulted in hospitalization, an indicator of severe infections. This study confirms  
32 the highly heterogeneous *S.aureus* population in Kenya. MSSA appear to have the potential of  
33 accumulating more toxin genes than MRSA. This co-occurrence of major toxin genes, some

34 associated with MRSA, highlights the potential risks of outbreaks of highly virulent MRSA  
35 infections which would pose treatment challenges.

36 **Keywords:** MRSA, MSSA, Kenya, virulence, whole genome

## 37 **Introduction**

38 *Staphylococcus aureus* is one of the leading causes of nosocomial infections but in recent years  
39 it has been increasingly associated with community acquired infections [1]. *S.aureus* causes a  
40 wide spectrum of diseases including bacteremia, pneumonia, urinary tract, skin and soft tissue  
41 infections (SSTI) [2,3]. Since the emergence of methicillin resistance (MRSA) in the 1940s  
42 epidemics initiated by successful MRSA clones have been observed [4,5] . For instance, USA 300,  
43 a highly virulent MRSA strain that emerged in the USA is currently associated with community  
44 outbreaks globally [6]. E-MRSA 15 which emerged in the UK has been linked to various hospital  
45 outbreaks [7]. Clonal success has been attributed to factors that enhance binding to host tissues  
46 and to the acquisition of virulence genes. For example, USA 300 has acquired the arginine  
47 catabolic mobile element, *sek* and *seq* virulence genes [8-10].

48 Strain typing is necessary for identification of emerging and outbreak associated strains. Multi  
49 locus sequence typing (MLST) and typing of the staphylococcal protein A (*spa*) gene have been  
50 widely used over the last 10 years to identify different strain types [11,12]. MLST examines  
51 differences in 7 housekeeping genes, assigns allele numbers and sequence types (ST) from the  
52 allelic profiles. These ST are then grouped into larger groups known as clonal complexes [13].  
53 Spa typing is based on the number, sequence and type of repeats in the hypervariable region of

54 protein A. For MRSA, additional typing of the staphylococcal cassette chromosome (SSC mec)  
55 which harbors the gene encoding methicillin resistance allows for additional discrimination  
56 between MRSA strains. Staphylococcal cassettes differ in gene arrangement and at least 11  
57 types have been reported [14-16].

58 The *S.aureus* genome bears a plethora of virulence determinants [17,18] some of which are  
59 encoded in the core genome while others are borne on accessory genomes such as plasmids,  
60 conjugative transposons, plasmids and cassettes [19]. Virulence factors are grouped according  
61 to function into adhesins, immune evasion proteins, toxins and pore forming proteins. Cell wall  
62 adhesins mediate binding to host tissues and biofilm formation e.g. clumping factor A and  
63 polysaccharide intracellular adhesin. Pore forming proteins include leukocidins (LukE, LukD, and  
64 PVL) and hemolysins (gamma, beta, alpha). *S.aureus* toxins have been associated with severe food  
65 poisoning (SEB, SEA) exfoliative skin conditions (ETA, ETB) and systemic shock (TSST-1) [20-24].  
66 The distribution of virulence genes has been shown to differ between strain types [25].

67 Previous Kenyan studies have largely focused on susceptibility profiles [26,27] of *S.aureus* and  
68 strain typing with limited testing for virulence determinants which can influence infection  
69 severity. Virulence genes reported in both *S.aureus* carriage and infection studies in Kenya  
70 include Panton Valentine leukocidin (*PVL*), Toxic shock syndrome (TSST1), exfoliative toxin A and  
71 enterotoxin A with a notably high prevalence of *PVL* reported [28,29]. These studies were limited  
72 to 4 healthcare institutions in close geographic proximity therefore there is limited information  
73 on the diversity and distribution of the *S.aureus* population across Kenya. This study sought to  
74 fill this gap by typing isolates from a wider geographical area and identifying the relationships

75 between Kenyan strains and known global strains using phylogeny. In addition, isolates were  
76 screened for a panel of known virulence genes and attempts made to correlate their presence  
77 with disease severity. By broadening our understanding of the *S.aureus* population in Kenya, this  
78 study provides baseline data for tracking emerging hypervirulent or outbreak associated strains  
79 of *S.aureus*.

## 80 **Materials and methods**

### 81 **Ethics**

82 This study was approved by the Walter Reed Army Institute of Research (#2089) and Kenya  
83 Medical Research Institute (#2767) IRBs.

### 84 **Bacterial isolates identification**

85 Clinical *S.aureus* isolates from patients enrolled in an ongoing surveillance study (WRAIR#2089,  
86 KEMRI#2767) in public hospitals in three counties (Kisumu, Kericho and Nairobi) in Kenya were  
87 analyzed for this study. *S.aureus* isolates were identified based on beta hemolysis on sheep  
88 blood agar plates, gram positive, clustered cocci by gram stain, catalase and coagulase positive  
89 phenotypes. Isolate identity was confirmed using the MALDI-TOF biotyper (Bruker Daltonics,  
90 Millerica, MA, USA) and using the GP card on the Vitek 2 platform (bioMérieux, Hazelwood, MO,  
91 USA). 19 isolates (4 MRSA, 15 MSSA) obtained between April 2015 and December 2015 were  
92 studied. A majority of the *S.aureus* isolates were from skin and soft tissue infections (18/19).  
93 84.2% were community acquired infections and 15.8% (3/19) were hospital acquired infections

94 as per the CDC classification (29). Outpatient infections were considered mild infections (10/19,  
95 52.6%) whereas inpatient infections (9/19, 47.3%) were considered severe infections, Table 1.

97

98 **Table 1: Clinical characteristics of the Kenyan *S.aureus* isolates in this study**

ISOLATE #	INFECTION TYPE <sup>d</sup>	MRSA/MSSA <sup>b</sup>	CA/HA <sup>c</sup>	IP/OP <sup>a</sup>	SEVERITY
SAKEN01	UTI	MRSA	CA	OP	Mild
SAKEN05	SSTI	MRSA	HA	IP	Severe
SAKEN06	SSTI	MRSA	CA	IP	Severe
SAKEN13	SSTI	MSSA	CA	OP	Mild
SAKEN14	SSTI	MSSA	HA	IP	Severe
SAKEN15	SSTI	MSSA	HA	IP	Severe
SAKEN16	SSTI	MSSA	CA	OP	Mild
SAKEN17	SSTI	MSSA	CA	OP	Mild
SAKEN18	SSTI	MSSA	CA	IP	Severe
SAKEN19	SSTI	MSSA	CA	IP	Severe
SAKEN20	SSTI	MSSA	CA	IP	Severe
SAKEN21	SSTI	MRSA	CA	IP	Severe
SAKEN22	SSTI	MSSA	CA	OP	Mild
SAKEN23	SSTI	MSSA	CA	IP	Severe
SAKEN24	SSTI	MSSA	CA	OP	Mild
SAKEN25	SSTI	MSSA	CA	OP	Mild
SAKEN26	SSTI	MSSA	CA	IP	Severe
SAKEN27	SSTI	MSSA	CA	OP	Mild
SAKEN29	SSTI	MSSA	CA	OP	Mild

99 Table legend: <sup>a</sup>IP-Inpatient, OP-Outpatient. Inpatient infections are considered severe infections in this  
100 study. <sup>b</sup>Methicillin Resistant *Staphylococcus aureus* (MRSA), Methicillin Susceptible *Staphylococcus aureus*.  
101 <sup>c</sup>CA – Community acquired infection, HA - Hospital associated infection. <sup>d</sup>SSTI – Skin and soft tissue  
102 infection, UTI – Urinary tract infection

## 103 **Isolate typing**

104 In-vitro spa typing was performed using conventional PCR [30] and Sanger sequencing. Contigs  
105 were assembled on CLC bio Main-Workbench (CLC bio, Aarhus, Denmark) and spa types  
106 analyzed using Ridom StaphType (Ridom GmbH, Münster, Germany). In-silico spa typing was  
107 done by analyzing assembled genomes on the online analysis pipeline  
108 <https://cge.cbs.dtu.dk/services/spatyper/>. Spa types obtained by both methods were compared.  
109 Sequences of isolates with novel spa repeats were submitted to the Ridom Spa Server for  
110 assignment of spa type.

111 MLST sequence type (ST) was determined in-vitro using published primers [11]. Gene sequences  
112 for each of the 7 loci were queried against the *S.aureus* database, allele numbers obtained and  
113 allelic profiles analyzed on <http://www.mlst.net/> [13] to assign ST. Isolates were grouped into  
114 clonal complexes using the BURST clustering algorithm available on <http://eburst.mlst.net/>,  
115 allowing a minimum of 6 identical loci for group definition. Sequences of novel ST were  
116 submitted to <https://pubmlst.org/saureus/> for ST assignment.

## 117 **Staphylococcal cassette chromosome typing**

118 Staphylococcal cassette types for the MRSA isolates were determined using previously published  
119 primers [14]. PCR products were visualized on agarose gels and SSC mec types determined  
120 based on different amplicon size.



## 121 **Whole genome sequencing and sequence analysis**

122 Genomic DNA was extracted from freshly cultured *S.aureus* isolates using ZR Fungal/Bacterial  
123 DNA MiniPrep Kit (Zymo research, California, United States). DNA concentrations were  
124 determined using the Qubit (Thermo Fisher Scientific, Massachusetts, United States) and 1ng of  
125 DNA used for library preparation with the Nextera XT kit (Illumina) as per manufacturer's  
126 instructions to generate 300bp paired end libraries. Libraries were sequenced on an Illumina  
127 MiSeq platform.

128 Raw reads were uploaded onto the Pathosystems Resource Integration Center (PATRIC 3.5)  
129 <https://www.PATRICbrc.org/> [31]. Genome assembly was carried out on PATRIC using the  
130 assembly pipeline 'miseq' as the assembly strategy which runs both Velvet [32] and SPAdes [33]  
131 algorithms and uses ARAST, an in-house scoring script. Annotation was performed using the  
132 RASTk pipeline [34] on PATRIC with 'Bacteria' as the domain and 'Staphylococcus aureus' as the  
133 taxonomy ID. Genome assemblies were uploaded onto NCBI under BioProject ID PRJNA481322.  
134 A summary of the genome characteristics can be found in Table 2.

## 135 **Bioinformatic analysis**

136 In-silico MLST was determined by analyzing the assembly files on both the PATRIC genome  
137 characterization pipeline and on MLST 1.8 <https://cge.cbs.dtu.dk/services/MLST/> [35] at the  
138 Centre for Genomic Epidemiology (CGE). Novel ST types were submitted to  
139 <https://pubmlst.org/saureus/> for typing [36].

140 To establish the relationships between the isolates of this study, MLST sequences for the query  
141 genomes were concatenated on FASconCAT [37] and maximum likelihood phylogeny inferred  
142 using PhyML 3.1 [38] utilizing the TMP model with gamma variation.

143 To infer relationships between genomes of this study and known global strains, phylogenetic  
144 analysis was performed on the 19 Kenyan *S.aureus* whole genomes and 26 global strains.  
145 Reference strains were selected to include at least one whole genome for all the sequence types  
146 identified and common global and regional strains. The reference strains used in the  
147 phylogenetic analysis are under supplementary information appendix S1, Table 1. To generate a  
148 whole genome sequence phylogeny on PATRIC, genome assemblies were analyzed on the  
149 phylogenetic tree building service. The tree building service filtered genome protein files and  
150 used BLAST to determine the best bi-directional hit, clustered protein files into homolog sets  
151 and filtered out homolog sets representing <80% of the genomes and eventually protein sets  
152 meeting the threshold were trimmed and aligned with MUSCLE [39]. Tree alignments were  
153 concatenated and a main tree built using RAxML [40]. Progressive tree refinement was  
154 employed for analysis of poorly refined sub – trees [31,41]. Phylogenetic trees generated from  
155 both the MLST loci and WGS (Figure 1-2) were compared.

156 To investigate the presence of known virulence genes, whole genomes of Kenyan isolates were  
157 screened for 56 virulence genes. Virulence genes were identified under the specialty genes on  
158 PATRIC with 'Virulence' as the search term. Targeted virulence genes and pathogenicity islands  
159 were identified across all isolates on PATRIC. To verify the virulence genes identified, contigs

160 were analyzed on VirulenceFinder 1.5 <https://cge.cbs.dtu.dk/services/VirulenceFinder/> [42]. The

161 virulence genes identified were then grouped according to function.

163 **Results and Discussion**

164 **Table 2: Genome characteristics of *S.aureus* isolates analyzed in this study**

<b>ISOLATE ID</b>	<b>GC CONTENT</b>	<b>GENOME SIZE(MB)</b>	<b>CONTIGS</b>	<b>coding sequences</b>	<b># of repeat regions</b>	<b># of tRNA present</b>	<b># of rRNA</b>
SAKEN01	32.65	3	66	2918	125	57	7
SAKEN05	32.7	2.97	83	2864	156	60	9
SAKEN06	32.7	2.97	73	2868	151	58	9
SAKEN13	32.7	2.7	73	2613	90	60	2
SAKEN14	32.7	2.73	68	2595	85	56	9
SAKEN15	32.7	2.71	50	2568	79	56	8
SAKEN16	32.78	2.71	91	2564	86	61	13
SAKEN17	32.79	2.74	103	2641	91	59	8
SAKEN18	32.77	2.6	98	2555	80	54	6
SAKEN19	32.74	2.8	44	2700	115	59	10
SAKEN20	32.76	2.79	96	2703	102	58	6
SAKEN21	32.69	2.9	83	2896	116	55	8
SAKEN22	32.73	2.76	56	2654	94	58	8
SAKEN23	32.69	2.72	58	2575	84	58	9
SAKEN24	32.82	2.72	145	2622	110	55	6
SAKEN25	32.87	2.69	183	2604	73	51	7
SAKEN26	32.74	2.83	89	2784	133	58	8
SAKEN27	32.8	2.74	52	2592	83	59	8
SAKEN29	32.69	2.83	111	2749	132	57	8

165 **Spa typing**

166 Study isolates typed into 8 distinct spa types (t005, t037, t064, t084, t233, t2029, t272, t355). t355  
167 was the dominant spa type (7/19; 36.8%). A single novel spa type, assigned t17826, was

168 reported. MRSA isolates typed to t007 (1) and t037 (3) while MSSA isolates showed great spa  
169 diversity with t005, t17826, t272, t13194 being represented by singletons. Discrepancies  
170 between in-vitro and in-silico spa typing were observed for SAKEN 01, (t007 vs t2029) and  
171 SAKEN 13, (t084 vs t233) (Table 2).

## 172 **MLST typing**

173 Multi locus sequence typing results were consistent on both CGE and PATRIC analysis platforms.  
174 There was great diversity among the isolates with the identification of 6 distinct CC  
175 (8,22,15,80,121,152) and 8 ST types (8, 15, 22,80,121,152,241, 1633) (Table 3). 2 novel STs  
176 assigned ST 4705 (CC 8, MRSA) and ST 4707 (CC 5, MSSA) by PubMLST [36] were reported. A  
177 majority of the isolates belonged to CC 152 (7/19) and CC 8 (5/19). ST 152 was the most  
178 dominant 6/19 (31%) with all isolates being MSSA geographically distributed across the 3  
179 counties. 4 of the STs were represented by singletons.

180 All the MRSA in this study belonged to CC 8. A majority of MRSA (3/4) in this study typed to ST  
181 241 and spa type t037. One isolate typed to the novel ST 4705 and spa type t007. This isolate  
182 only differed from the other isolates in the acetyl coenzyme A acetyltransferase (*yqiI*) loci. In-  
183 vitro SSC mec typing identified 2 SSC mec types with a majority of the MRSA typing to SSC mec  
184 IVc (3/4). The remaining isolate with the novel ST 4705, typed to SSC mec II.

186

187 **Table 3: Table showing STs, CCs of Kenyan *S.aureus* isolates**

Isolate	Clonal complex	MLST type (ST)	SPA type (in-vitro)	SPA type (in- silico)	MRSA/MSSA	REGION <sup>a</sup>
SAKEN01	8	4705*	<b>t007</b>	<b>t2029</b>	<b>MRSA</b>	KDH
SAKEN05	8	241	t037	t037	<b>MRSA</b>	KDH
SAKEN06	8	241	t037	t037	<b>MRSA</b>	KDH
SAKEN21	8	241	t037	t037	<b>MRSA</b>	KDH
SAKEN19	8	8	t064	t064	MSSA	KDH
SAKEN26	8	8	t064	t064	MSSA	KDH
SAKEN16	152	152	t355	t355	MSSA	KCH
SAKEN18	152	152	t355	t355	MSSA	KCH
SAKEN20	152	152	t355	t355	MSSA	KDH
SAKEN14	152	152	t355	t355	MSSA	NRB
SAKEN22	152	152	t355	t355	MSSA	NRB
SAKEN23	152	152	t355	t355	MSSA	NRB
SAKEN15	152	1633	t355	t355	MSSA	NRB
SAKEN24	22	22	t005	t005	MSSA	KCH
SAKEN13	22	22	<b>t084</b>	<b>t233</b>	MSSA	NRB
SAKEN25	5	4707*	t17826*	unknown	MSSA	KDH
SAKEN17	15	15	t084	t084	MSSA	KCH
SAKEN27	80	80	t13194	t13194	MSSA	KCH
SAKEN29	121	121	t272	t272	MSSA	KDH

188 Table legend. \* denotes novel ST and spa types reported by this study. <sup>a</sup> Three regions in Kenya where the

189 isolates were obtained from: KCH- Kericho, KDH – Kisumu County, NRB- Nairobi County

190 Each region showed a unique genetic fingerprint with CC 8 (ST 8 and ST 241) detected in Kisumu  
191 County only while CC 152 showed a wide geographic distribution. Kericho County showed  
192 greatest heterogeneity in CC/ST.

## 193 **Phylogenetic analysis**

194 Maximum likelihood phylogenies reconstructed on PATRIC and on PhyML gave similar tree  
195 topologies with isolates clustering according to sequence types and/or clonal complexes Fig 1-2.

196

197 **Fig 1: Dendrogram of an MLST phylogeny of Kenyan *S.aureus* isolates.** Maximum likelihood was phylogeny  
198 inferred using a custom model and 100 boot strap replicates. ST and CC associated with MRSA isolates of this study  
199 are depicted in bold font.

200 The phylogenetic tree inferred using MLST depicted a multifurcating tree with distribution of  
201 MRSA and MSSA isolates in distinct clusters. ST 121 was ancestral to ST 152 while ST 80 was  
202 ancestral to ST 22. CC 8 cluster encompassed all MRSA isolates (ST 241, ST 4705) and ST 8  
203 MSSA isolates. The MRSA clade (ST 241) and MSSA (ST 8) clade are sister groups sharing a  
204 recent common ancestor. The MRSA isolate (SAKEN 01) which has a novel ST grouped within the  
205 MRSA clade.

206

207 **Fig 2: Dendrogram showing whole genome phylogeny of Kenyan isolates** (SAKEN prefix and in blue)  
208 relative to known global strains. MRSA isolates are depicted in green. Reference isolates are in black with  
209 a country/regional prefix. Phylogeny inferred using maximum likelihood with progressive refinement on  
210 PATRIC.

211 Whole genome phylogeny with global strains showed Kenyan ST 152 isolates clustered with the  
212 asymptomatic Malian BB155 strain. MRSA in this study were closely related to TW20 strain 582  
213 (ST 239) which is a successful HA MRSA lineage in the United Kingdom [7]. Kenyan ST8 MSSA  
214 isolates grouped closely with the highly virulent Newman strain [43]. ST/CC 22 was represented  
215 by 2 MSSA isolates which clustered with EMRSA-15 (MRSA-15), associated with widespread  
216 outbreak of HA-MRSA in Europe [44]. It was also observed that Kenyan strains in this study were  
217 closely related to MRSA and MSSA isolates from Tanzania and Sudan respectively. Tanzanian  
218 MRSA isolates clustered with the Kenyan ST 8 MSSA isolates and Kenyan ST 15 MSSA.

## 219 **Virulence genes screen.**

220 The 19 *S.aureus* genomes were screened in-silico for a panel of 56 known virulence genes (19  
221 adhesins, 9 hemolysins, 5 immune evasion proteins, 6 exo-enzymes and 19 toxins). Of the 19  
222 adhesins investigated, 10 were identified in all the isolates (ClfA, ClfB, FnbA, EbpS, IcaA,  
223 Eap/Map, SasA, SacC, SasF, and SasH) (S2 table). Since these adhesins are ubiquitously  
224 expressed as they are essential for successful host invasion by facilitating binding to extra  
225 cellular matrix proteins [45], they are not discussed further in this paper. Efb (Extracellular  
226 Fibrinogen – binding protein) was detected in 84% of the isolates. 18/19 of the isolates bore  $\geq 2$   
227 of the Ser-Asp rich fibrinogen binding proteins whereas collagen binding protein (cna) was  
228 present in 63% of the isolates. It was observed that the presence of the adhesin genes was  
229 loosely correlated with the clonal complex. For example, isolates lacking Efb (3) belonged to CC  
230 152 while all CC 152 isolates lacked SdrC, SasG cell wall adhesion protein and CC8 isolates  
231 lacked the SasK adhesion protein.



232 A majority of the hemolysins screened for (6/9) were identified in the isolates (gamma A, B, C,  
233 delta, alpha hemolysin and PVL). 63% (12/19) of the isolates had  $\geq 5$  of the hemolysins screened  
234 for with Leukocidin D and E being present in 94% and 89% of the isolates respectively.  
235 Hemolysins damage the host cell plasma membrane which is a critical process during *S.aureus*  
236 infection and disease progression [46]. Of note, all isolates typing to CC 152 lacked beta  
237 hemolysin whereas the bi-component protein Panton Valentine leukocidin was detected in all  
238 the Kenyan isolates.

239 Of the 19 toxins screened for 9 were detected; TSST-1, ETA/ETB, SEB, SEG, SEH, SEK, SEL, SEQ,  
240 and SPEG. Toxins were found to be randomly distributed among the CCs. TSST-1 was detected  
241 in 3/19 of the isolates obtained from various soft tissue infections (burns, abscesses and cuts).  
242 All the isolates with TSST-1 were MSSA of which 2/3 were from hospitalized patients an indicator  
243 of severe infection. Staphylococcal pathogenicity island 1 (SaP1) was detected in all isolates with  
244 TSST-1 (S2 Table 1). SEB was present in 2/19 of the isolates obtained from hospitalized patients  
245 presenting with a burn and an abscess. SEL, SEG enterotoxins and ETA/ETB exfoliative toxins  
246 were each detected in 2/19 of the isolates. Isolates with ETA and ETB exfoliative toxins were  
247 obtained from outpatient subjects with mild abscess and skin lesion infections. Staphylococcal  
248 pathogenicity island 2 (SaPln2) was detected in isolates with one or both exfoliative toxins.

249 Some toxin combinations were observed to always co-occur: SEK+SEQ, SEB+SEL, TSST-1+SEG,  
250 TSST-1+SEB+SEL, TSST-1+SEB+SEK+SEL. The SEK+SEQ combination was present in all the  
251 MRSA isolates in this study a majority (3/4) of which were from hospitalized subjects. TSST-1  
252 detected in this study was from both hospitalized (2) and outpatient (1) subjects. In the

253 hospitalized patients TSST-1 co-occurred with SEB and SEL while in an outpatient TSST-1 co-  
254 occurred with SEG suggesting that TSST-1+SEB+SEL combination could cause severe infections.  
255 Comparison of the number of toxin genes between MRSA and MSSA isolates revealed that all  
256 Kenyan MRSA isolates only had 2 toxins whereas MSSA isolates had between 0-5 toxins. The  
257 isolate with the highest number of toxins (5) was an MSSA isolate obtained from a hospitalized  
258 burn patient. This MSSA isolate was in the same CC as the MRSA isolates suggesting the  
259 potential risk of emergence of CC 8 strains that are both highly virulent and difficult to treat. A  
260 majority (2/3) of the hospital acquired infections had no toxin genes detected in their genomes  
261 with the remaining isolate having 2 toxin genes.

## 262 **Discussion**

263 This study typed *S.aureus* isolates from widespread geographical areas in Kenya and inferred  
264 phylogenetic relationships between the isolates and known global and regional strains. To  
265 understand the virulence potential of the Kenyan strains, virulence genes in the Kenyan strains  
266 were identified and the presence of these genes investigated in relation to CC and clinical  
267 presentation.

268 Strain typing revealed 8 STs and 8 spa types among the 19 Kenyan isolates confirming the great  
269 heterogeneity previously described among *S.aureus* both regionally and globally [28,29,47-49].  
270 The isolates grouped into 6 distinct clonal complexes with CC 152 being most dominant and  
271 widely distributed across the three counties. CCs 8 and 22 known to harbor MRSA globally were

272 also reported. All MRSA isolates in this study typed to CC 8 while MSSA isolates showed  
273 heterogeneous distribution across a number of CCs.

274 Globally, MRSA strains have been shown to belong to 3 major CC: CC 8, CC 5 and CC 22 [50].  
275 For instance, EMRSA15/UK that is responsible for hospital acquired MRSA infections in the UK  
276 types to CC 22. In this study CC 22 was comprised of only MSSA isolates whereas CC 8 was  
277 composed of both MSSA and MRSA isolates. Previous Kenyan studies carried out in hospitals in  
278 Nairobi [47] and Thika [28] identified CC 5 as the predominant circulating MRSA clone within  
279 hospitals in Nairobi and its environs. MRSA isolates from hospitals in Nairobi were  
280 predominantly ST 241, t037 while ST 239, t037 prevailed in the hospital in Thika. In this study, all  
281 MRSA isolates typed to ST 241, t037; similar to that reported by Omuse et al [47] in 3 hospitals  
282 in Nairobi. All MRSA isolates in this study (CC 8, ST 241 and novel ST 4705) were obtained from  
283 Western Kenya which is situated 300km from Nairobi. This identification of ST 241 MRSA strains  
284 in both Western Kenya and Nairobi suggests a widespread geographical distribution of this  
285 MRSA strain in Kenya. Schaumburg et al [48] reported ST 241 MRSA clone to be widespread in  
286 Africa though with varying SSCmec types; Senegal (SSCmec III), Tunisia (SSCmec III), Niger  
287 (SSCmec III and V) Nigeria (SSCmec III and IV) and Algeria (SSCmec III) [51]. A majority of the  
288 MRSA isolates of this study bore SSCmev IV similar to that reported in Nigeria.

289 Spa typing showed a greater discriminatory power than ST with multiple spa types belonging to  
290 the same STs. Discrepancies between in-vitro and in-silico spa types (t007 vs t2029 and t084 vs  
291 t2330) was due to shorter repeat sequences in-silico. The shorter repeats observed in silico can  
292 be attributed to filtering out of low quality reads in the sequence analysis pipeline. Thus, we

293 recommend the use of conventional in-vitro spa typing as a surveillance tool more so in regions  
294 with great *S.aureus* population diversity.

295 Inferred phylogeny of the Kenyan isolates showed distinctive clustering by CC. CC 8 cluster  
296 composed of two clades ST 8 (MSSA) and ST 241 (MRSA) with the two clades sharing a recent  
297 common ancestor. Studies have shown that MSSA isolates of CC 8 act as reservoirs for MRSA  
298 pending acquisition of the staphylococcal cassette [52,53].

299 Relationships between isolates of this study and known global strains using whole genome  
300 phylogeny revealed close clustering of MRSA strains in this study with the well-known TW20  
301 strain 582, which is a successful HA MRSA clone which originated from London in the UK [7].  
302 TW20 is a hospital associated outbreak MRSA strain known for its high transmissibility and  
303 multi-drug resistant properties due to a plethora of resistance genes carried on mobile elements  
304 [54]. TW20 and the MRSA strains in this study both type to CC 8. CC 8 MRSA strains have been  
305 linked to community acquired infections and is the predominant MRSA strain in this study. This  
306 clonal complex encompasses well known strains such as USA 300 which is a lineage linked to the  
307 acquisition of SSCmec IV, PVL and SEQ and SEK genes [10,55]. EMRSA-15 strain belongs to the  
308 same clade as Kenyan ST 22 MSSA isolates. EMRSA-15 is a well-known strain associated with the  
309 widespread outbreak of HA-MRSA in Europe [44] and MSSA isolates typing to ST 22 have been  
310 identified as the MSSA reservoir from which EMRSA-15 emerged [56]. ST 152, the dominant  
311 Kenyan strain is related to both carriage strains in Mali [57] and pathogenic strains in Europe  
312 [58-60].

313 Bacterial virulence factors are key for successful host colonization and infection. Adhesins  
314 facilitate successful binding to the host extra cellular matrix promoting subsequent biofilm  
315 formation. Worth noting is that ST 152 isolates in this study lacked some adhesins (SasG and  
316 SdrC) and hemolysin (beta) genes. SasG surface protein mediates successful bacterial adhesion  
317 to squamous epithelial cells of the nostrils and promotes biofilm formation. Expression of SasG  
318 has been shown to mask the effect of adhesins binding to ligands such as fibrinogen and  
319 fibronectin [61]. Both Malian BB155 and Kenyan ST 152 isolate genomes lack SasG which we  
320 speculate could promote successful host colonization and provide a fitness advantage. All CC8  
321 isolates lacked SasK cell wall adhesion, the significance of this putative adhesin is yet to be  
322 determined [62]. Previous studies by McCarthy and Lindsay [25] noted similar observation of  
323 variation in surface adhesions between clonal complexes and lineages.

324 PVL, a bi-component leukocidin causing destruction of leukocytes and tissue necrosis, was  
325 widespread across all isolates in this study. PVL positive MRSA clones were identified in this  
326 study and in other Kenyan and African studies [27,29,63,64]. A Nigerian study pointed out the  
327 possibility of emergence of PVL positive MRSA clones as a result of the co-existence of MRSA  
328 clones and PVL positive MSSA [65]. This study reports extensive (100%) PVL presence in Kenyan  
329 MSSA and MRSA isolates supporting this hypothesis.

330 Staphylococcal toxin (SEB) is a super antigenic toxin associated with food poisoning, non-  
331 menstrual toxic shock syndrome, dermatitis and asthma [20]. Kenyan isolates bearing SEB typed  
332 to CC 8 and CC 152 consistent with previous observations where SEB was identified most often  
333 in CC8 isolates in New York [66]. A Taiwanese study identified SEB to be the cause of

334 Staphylococcal scarlet fever [67]. In this study, SEB was detected in MSSA isolates obtained from  
335 a burn and from an abscess. In both cases these were inpatients indicative of severe infections.

336 Toxic shock syndrome toxin (TSST-1) was initially reported as the cause of menstrual toxic shock  
337 but over the years non-menstrual toxic shock syndrome has been reported [68-70]. TSST-1 is  
338 encoded by *tstH* gene borne on the staphylococcal pathogenicity island 1 [71]. SaP1 was  
339 identified in all the isolates positive for TSST-1. TSST-1 was observed to co-occur with up to 4  
340 staphylococcal enterotoxins and most often with SEB and SEL. The presence of TSST-1 and SEB  
341 toxins in isolates obtained from severe infections resulting in hospital admissions suggests the  
342 severity of this toxin combination.

343 SEK and SEQ have been linked to a number of food poisoning cases [24,72] . Among the Kenyan  
344 isolates these enterotoxins were associated with SSTIs. SEQ+SEK combination was consistently  
345 observed in Kenyan MRSA isolates. SEQ and SEK toxin genes were reported to significantly co-  
346 occur in Chinese MRSA isolates [73]. These toxins co-occur on genomic islands and have been  
347 associated with the HA SSC *mec II* clone [21,66]. CC 8 isolates have been reported to  
348 consistently bear SEQ, SEL and SEK toxin genes [19]. This pattern of distribution was observed  
349 among isolates in this study.

350 Of the 4 MRSA isolates 3 were MRSA ST 241, *mec IVc* and the remaining isolate was the novel  
351 MRSA ST 4705, *mec II*. Worth noting was the difference between the SSC *mec* cassettes present  
352 in the isolates and that the MRSA ST 4705, *mec II* was the sole MRSA isolate associated with a  
353 community acquired UTI infection.

354 It was also observed that a majority of MRSA isolates from this study were community acquired  
355 and bore PVL, enterotoxin Q and K toxin genes. Studies by Voyich et al [74] have suggested the  
356 highly virulent nature of community acquired MRSA in comparison to hospital acquired strains a  
357 hypothesis that is not supported by this study which showed no differences in numbers and  
358 types of virulence genes between the hospital acquired and the community acquired strains.

## 359 **Conclusion**

360 Despite the low number of isolates analyzed, this study provides a glimpse into the diversity and  
361 distribution of Kenyan MSSA and MRSA isolates and their relatedness to global strains. The  
362 study highlighted the potential impact of particular toxin combinations on clinical severity and  
363 provided evidence that co-occurrence of methicillin resistance and virulence genes could  
364 portend the emergence of highly virulent MRSA infections. There is demonstrated need for  
365 continued trend monitoring through surveillance which will continue as part of this ongoing  
366 surveillance program.

## 367 **Acknowledgements**

368 We thank Dr John Waitumbi, Kimita Gathii and Luiser Ingasia from Kisumu basic Science Lab for  
369 sequencing support, Ann Wattam (PATRIC) and Frances Coll (Sanger Institute) for assistance in  
370 WGS analysis and phylogeny. We are grateful to all the study subjects in the participating  
371 healthcare facilities. This work has been published with the permission of the Director KEMRI.  
372 Material has been reviewed by the Walter Reed Army Institute of Research. There is no  
373 objection to its publication. The opinions or assertions contained herein are the private views of

374 the author, and are not to be construed as official, or as reflecting true views of the Department  
375 of the Army or the Department of Defense. The investigators have adhered to the policies for  
376 protection of human subjects as prescribed in AR 70–25.

## 377 **Author contributions**

378 Conceptualization: LM  
379 Data Curation: LM, CK  
380 Formal analysis: CK, JN, LM  
381 Funding Acquisition: LM  
382 Investigation: CK, JN, DM, VO, SW  
383 Methodology: LM  
384 Supervision: VO, LM, WS  
385 Validation: LM, CK  
386 Visualization: LM, CK  
387 Writing- Original draft preparation: CK, LM  
388 Writing- Review and Editing: JN, CK, VO, SW, DM, WS, LM

## 389 **REFERENCES**

- 390 1. Klein E, Smith DL, Laxminarayan R (2007) Hospitalizations and Deaths Caused by Methicillin-  
391 Resistant *Staphylococcus aureus*, United States, 1999–2005. *Emerg Infect Dis* 13: 1840-  
392 1846.
- 393 2. (2011) *Staphylococcus aureus* in Healthcare Settings. Centers for Disease Control and  
394 Prevention.
- 395 3. Lowy FD (1998) *Staphylococcus aureus* infections. *N Engl J Med* 339: 520-532.
- 396 4. Chambers HF, DeLeo FR (2009) Waves of Resistance: *Staphylococcus aureus* in the Antibiotic  
397 Era. *Nat Rev Microbiol* 7: 629-641.
- 398 5. Mediavilla JR, Chen L, Mathema B, Kreiswirth BN (2012) Global epidemiology of community-  
399 associated methicillin resistant *Staphylococcus aureus* (CA-MRSA). *Curr Opin Microbiol*  
400 15: 588-595.



- 401 6. Nimmo GR (2012) USA300 abroad: global spread of a virulent strain of community-associated  
402 methicillin-resistant *Staphylococcus aureus*. *Clin Microbiol Infect* 18: 725-734.
- 403 7. Edgeworth JD, Yadegarfar G, Pathak S, Batra R, Cockfield JD, et al. (2007) An outbreak in an  
404 intensive care unit of a strain of methicillin-resistant *Staphylococcus aureus* sequence  
405 type 239 associated with an increased rate of vascular access device-related bacteremia.  
406 *Clin Infect Dis* 44: 493-501.
- 407 8. Tenover FC, McDougal LK, Goering RV, Killgore G, Projan SJ, et al. (2006) Characterization of a  
408 strain of community-associated methicillin-resistant *Staphylococcus aureus* widely  
409 disseminated in the United States. *J Clin Microbiol* 44: 108-118.
- 410 9. Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, et al. (2004) Complete genomes of two  
411 clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and  
412 drug resistance. *Proc Natl Acad Sci U S A* 101: 9786-9791.
- 413 10. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, et al. (2006) Complete genome sequence of  
414 USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus*  
415 *aureus*. *Lancet* 367: 731-739.
- 416 11. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000) Multilocus sequence typing for  
417 characterization of methicillin-resistant and methicillin-susceptible clones of  
418 *Staphylococcus aureus*. *J Clin Microbiol* 38: 1008-1015.
- 419 12. Strommenger B, Bräulke C, Heuck D, Schmidt C, Pasemann B, et al. (2008) *spa* Typing of  
420 *Staphylococcus aureus* as a frontline tool in epidemiological typing. *J Clin Microbiol* 46:  
421 574-581.
- 422 13. Aanensen DM, Spratt BG (2005) The multilocus sequence typing network: *mlst.net*. *Nucleic*  
423 *Acids Res* 33: W728-733.
- 424 14. Zhang K, McClure JA, Elsayed S, Louie T, Conly JM (2005) Novel multiplex PCR assay for  
425 characterization and concomitant subtyping of staphylococcal cassette chromosome  
426 *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 43: 5026-  
427 5033.
- 428 15. Ito T, Katayama Y, Asada K, Mori N, Tsutsumimoto K, et al. (2001) Structural comparison of  
429 three types of staphylococcal cassette chromosome *mec* integrated in the chromosome  
430 in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 45: 1323-  
431 1336.
- 432 16. (2009) Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): guidelines for  
433 reporting novel SCC*mec* elements. *Antimicrob Agents Chemother* 53: 4961-4967.
- 434 17. Otto M (2013) Community-associated MRSA: what makes them special? *Int J Med Microbiol*  
435 303: 324-330.
- 436 18. Diep BA, Carleton HA, Chang RF, Sensabaugh GF, Perdreau-Remington F (2006) Roles of 34  
437 virulence genes in the evolution of hospital- and community-associated strains of  
438 methicillin-resistant *Staphylococcus aureus*. *J Infect Dis* 193: 1495-1503.
- 439 19. Li M, Diep BA, Villaruz AE, Braughton KR, Jiang X, et al. (2009) Evolution of virulence in  
440 epidemic community-associated methicillin-resistant *Staphylococcus aureus*. *Proc Natl*  
441 *Acad Sci U S A* 106: 5883-5888.
- 442 20. Fries BC, Varshney AK (2013) Bacterial Toxins—Staphylococcal Enterotoxin B. *Microbiol*  
443 *Spectr* 1.

- 444 21. Cheng J, Wang Y, Cao Y, Yan W, Niu X, et al. (2016) The Distribution of 18 Enterotoxin and  
445 Enterotoxin-Like Genes in *Staphylococcus aureus* Strains from Different Sources in East  
446 China. *Foodborne Pathog Dis* 13: 171-176.
- 447 22. Adesiyun AA, Lenz W, Schaal KP (1991) Exfoliative toxin production by *Staphylococcus*  
448 *aureus* strains isolated from animals and human beings in Nigeria. *Microbiologica* 14:  
449 357-362.
- 450 23. Bukowski M, Wladyka B, Dubin G (2010) Exfoliative Toxins of *Staphylococcus aureus*. *Toxins*  
451 (Basel) 2: 1148-1165.
- 452 24. Ikeda T, Tamate N, Yamaguchi K, Makino S (2005) Mass Outbreak of Food Poisoning Disease  
453 Caused by Small Amounts of Staphylococcal Enterotoxins A and H. *Appl Environ*  
454 *Microbiol* 71: 2793-2795.
- 455 25. McCarthy AJ, Lindsay JA (2010) Genetic variation in *Staphylococcus aureus* surface and  
456 immune evasion genes is lineage associated: implications for vaccine design and host-  
457 pathogen interactions. *BMC Microbiol* 10: 173.
- 458 26. Omuse G, Kabera B, Revathi G (2014) Low prevalence of methicillin resistant *Staphylococcus*  
459 *aureus* as determined by an automated identification system in two private hospitals in  
460 Nairobi, Kenya: a cross sectional study. *BMC Infect Dis* 14: 669.
- 461 27. Maina EK, Kiiyukia C, Wamae CN, Waiyaki PG, Kariuki S (2013) Characterization of methicillin-  
462 resistant *Staphylococcus aureus* from skin and soft tissue infections in patients in  
463 Nairobi, Kenya. *Int J Infect Dis* 17: e115-119.
- 464 28. Aiken AM, Mutuku IM, Sabat AJ, Akkerboom V, Mwangi J, et al. (2014) Carriage of  
465 *Staphylococcus aureus* in Thika Level 5 Hospital, Kenya: a cross-sectional study.  
466 *Antimicrob Resist Infect Control* 3: 22.
- 467 29. Omuse G, Shivachi P, Kariuki S, Revathi G (2013) Prevalence of Panton Valentine Leukocidin  
468 in Carriage and Infective Strains of *Staphylococcus aureus* at a Referral Hospital in Kenya.  
469 *Open Journal of Medical Microbiology* Vol.03No.01: 7.
- 470 30. Shopsis B, Gomez M, Montgomery SO, Smith DH, Waddington M, et al. (1999) Evaluation of  
471 protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus*  
472 *aureus* strains. *J Clin Microbiol* 37: 3556-3563.
- 473 31. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, et al. (2014) PATRIC, the bacterial  
474 bioinformatics database and analysis resource. *Nucleic Acids Res* 42: D581-591.
- 475 32. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de  
476 Bruijn graphs. *Genome Res* 18: 821-829.
- 477 33. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. (2012) SPAdes: A New  
478 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput*  
479 *Biol* 19: 455-477.
- 480 34. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, et al. (2015) RASTtk: a modular and  
481 extensible implementation of the RAST algorithm for building custom annotation  
482 pipelines and annotating batches of genomes. *Sci Rep* 5: 8365.
- 483 35. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, et al. (2012) Multilocus sequence  
484 typing of total-genome-sequenced bacteria. *J Clin Microbiol* 50: 1355-1361.
- 485 36. Jolley KA, Maiden MC (2010) BIGSdb: Scalable analysis of bacterial genome variation at the  
486 population level. *BMC Bioinformatics* 11: 595.

- 487 37. Kuck P, Longo GC (2014) FASconCAT-G: extensive functions for multiple sequence alignment  
488 preparations concerning phylogenetic studies. *Front Zool* 11: 81.
- 489 38. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and  
490 methods to estimate maximum-likelihood phylogenies: assessing the performance of  
491 PhyML 3.0. *Syst Biol* 59: 307-321.
- 492 39. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high  
493 throughput. *Nucleic Acids Res* 32: 1792-1797.
- 494 40. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
495 large phylogenies. *Bioinformatics* 30: 1312-1313.
- 496 41. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, et al. (2017) Improvements to PATRIC, the  
497 all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 45:  
498 D535-D542.
- 499 42. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, et al. (2014) Real-time whole-genome  
500 sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic  
501 *Escherichia coli*. *J Clin Microbiol* 52: 1501-1510.
- 502 43. Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K (2008) Genome sequence of  
503 *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal  
504 genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol*  
505 190: 300-310.
- 506 44. Johnson AP, Aucken HM, Cavendish S, Ganner M, Wale MC, et al. (2001) Dominance of  
507 EMRSA-15 and -16 among MRSA causing nosocomial bacteraemia in the UK: analysis of  
508 isolates from the European Antimicrobial Resistance Surveillance System (EARSS). *J*  
509 *Antimicrob Chemother* 48: 143-144.
- 510 45. Hauck CR, Agerer F, Muenzner P, Schmitter T (2006) Cellular adhesion molecules as targets  
511 for bacterial infection. *Eur J Cell Biol* 85: 235-242.
- 512 46. Vandenesch F, Lina G, Henry T (2012) *Staphylococcus aureus* hemolysins, bi-component  
513 leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging  
514 virulence factors? *Front Cell Infect Microbiol* 2: 12.
- 515 47. Omuse G, Van Zyl KN, Hoek K, Abdulgader S, Kariuki S, et al. (2016) Molecular  
516 characterization of *Staphylococcus aureus* isolates from various healthcare institutions in  
517 Nairobi, Kenya: a cross sectional study. *Ann Clin Microbiol Antimicrob* 15: 51.
- 518 48. Schaumburg F, Alabi AS, Peters G, Becker K (2014) New epidemiology of *Staphylococcus*  
519 *aureus* infection in Africa. *Clin Microbiol Infect* 20: 589-596.
- 520 49. Rolo J, Miragaia M, Turlej-Rogacka A, Empel J, Bouchami O, et al. (2012) High genetic  
521 diversity among community-associated *Staphylococcus aureus* in Europe: results from a  
522 multicenter study. *PLoS One* 7: e34768.
- 523 50. Aanensen DM, Feil EJ, Holden MT, Dordel J, Yeats CA, et al. (2016) Whole-Genome  
524 Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of  
525 Invasive *Staphylococcus aureus* in Europe. *MBio* 7.
- 526 51. Abdulgader SM, Shittu AO, Nicol MP, Kaba M (2015) Molecular epidemiology of Methicillin-  
527 resistant *Staphylococcus aureus* in Africa: a systematic review. *Front Microbiol* 6: 348.

- 528 52. Driebe EM, Sahl JW, Roe C, Bowers JR, Schupp JM, et al. (2015) Using Whole Genome  
529 Analysis to Examine Recombination across Diverse Sequence Types of *Staphylococcus*  
530 *aureus*. PLoS One 10: e0130955.
- 531 53. Strauss L, Stegger M, Akpaka PE, Alabi A, Breurec S, et al. (2017) Origin, evolution, and global  
532 transmission of community-acquired *Staphylococcus aureus* ST8. Proc Natl Acad Sci U S  
533 A 114: E10596-E10604.
- 534 54. Holden MT, Lindsay JA, Corton C, Quail MA, Cockfield JD, et al. (2010) Genome sequence of a  
535 recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant  
536 of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). J Bacteriol 192:  
537 888-892.
- 538 55. Tristan A, Ferry T, Durand G, Dauwalder O, Bes M, et al. (2007) Virulence determinants in  
539 community and hospital methicillin-resistant *Staphylococcus aureus*. J Hosp Infect 65  
540 Suppl 2: 105-109.
- 541 56. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, et al. (2013) A genomic portrait of the  
542 emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus*  
543 pandemic. Genome Res 23: 653-664.
- 544 57. Ruimy R, Maiga A, Armand-Lefevre L, Maiga I, Diallo A, et al. (2008) The carriage population  
545 of *Staphylococcus aureus* from Mali is composed of a combination of pandemic clones  
546 and the divergent Panton-Valentine leukocidin-positive genotype ST152. J Bacteriol 190:  
547 3962-3968.
- 548 58. Perez-Roth E, Alcoba-Florez J, Lopez-Aguilar C, Gutierrez-Gonzalez I, Rivero-Perez B, et al.  
549 (2010) Familial furunculosis associated with community-acquired leukocidin-positive  
550 methicillin-susceptible *Staphylococcus aureus* ST152. J Clin Microbiol 48: 329-332.
- 551 59. Monecke S, Slickers P, Ellington MJ, Kearns AM, Ehricht R (2007) High diversity of Panton-  
552 Valentine leukocidin-positive, methicillin-susceptible isolates of *Staphylococcus aureus*  
553 and implications for the evolution of community-associated methicillin-resistant *S.*  
554 *aureus*. Clin Microbiol Infect 13: 1157-1164.
- 555 60. Muller-Premru M, Strommenger B, Alikadic N, Witte W, Friedrich AW, et al. (2005) New  
556 strains of community-acquired methicillin-resistant *Staphylococcus aureus* with Panton-  
557 Valentine leukocidin causing an outbreak of severe soft tissue infection in a football  
558 team. Eur J Clin Microbiol Infect Dis 24: 848-850.
- 559 61. Corrigan RM, Rigby D, Handley P, Foster TJ (2007) The role of *Staphylococcus aureus* surface  
560 protein SasG in adherence and biofilm formation. Microbiology 153: 2435-2446.
- 561 62. Paharik AE, Horswill AR (2016) The Staphylococcal Biofilm: Adhesins, Regulation, and Host  
562 Response. Microbiol Spectr 4.
- 563 63. Ramdani-Bouguessa N, Bes M, Meugnier H, Forey F, Reverdy ME, et al. (2006) Detection of  
564 methicillin-resistant *Staphylococcus aureus* strains resistant to multiple antibiotics and  
565 carrying the Panton-Valentine leukocidin genes in an Algiers hospital. Antimicrob Agents  
566 Chemother 50: 1083-1085.
- 567 64. Mbogori C, Muigai A, Kariuki S (2013) Detection and Characterization of Methicillin Resistant  
568 *Staphylococcus aureus* from Toilet and Classroom Door Handles in Selected Secondary  
569 Schools in Nairobi County. Open Journal of Medical Microbiology Vol.03No.04: 5.

- 570 65. Okon KO, Basset P, Uba A, Lin J, Oyawoye B, et al. (2009) Cooccurrence of predominant  
571 Panton-Valentine leukocidin-positive sequence type (ST) 152 and multidrug-resistant ST  
572 241 *Staphylococcus aureus* clones in Nigerian hospitals. *J Clin Microbiol* 47: 3000-3003.
- 573 66. Varshney AK, Mediavilla JR, Robiou N, Guh A, Wang X, et al. (2009) Diverse enterotoxin gene  
574 profiles among clonal complexes of *Staphylococcus aureus* isolates from the Bronx, New  
575 York. *Appl Environ Microbiol* 75: 6839-6849.
- 576 67. Wang CC, Lo WT, Hsu CF, Chu ML (2004) Enterotoxin B is the predominant toxin involved in  
577 staphylococcal scarlet fever in Taiwan. *Clin Infect Dis* 38: 1498-1502.
- 578 68. Takahashi N (2003) Neonatal toxic shock syndrome-like exanthematous disease (NTED).  
579 *Pediatr Int* 45: 233-237.
- 580 69. Davis JP, Chesney PJ, Wand PJ, LaVenture M (1980) Toxic-shock syndrome: epidemiologic  
581 features, recurrence, risk factors, and prevention. *N Engl J Med* 303: 1429-1435.
- 582 70. Shands KN, Schmid GP, Dan BB, Blum D, Guidotti RJ, et al. (1980) Toxic-shock syndrome in  
583 menstruating women: association with tampon use and *Staphylococcus aureus* and  
584 clinical features in 52 cases. *N Engl J Med* 303: 1436-1442.
- 585 71. Dinges MM, Orwin PM, Schlievert PM (2000) Exotoxins of *Staphylococcus aureus*. *Clin*  
586 *Microbiol Rev* 13: 16-34.
- 587 72. Chen TR, Chiou CS, Tsen HY (2004) Use of novel PCR primers specific to the genes of  
588 staphylococcal enterotoxin G, H, I for the survey of *Staphylococcus aureus* strains  
589 isolated from food-poisoning cases and food samples in Taiwan. *Int J Food Microbiol* 92:  
590 189-197.
- 591 73. Bing-Shao Liang Y-MH, Yin-Shuang Chen, Hui Dong, Jia-Liang Mai, Yong-Qiang Xie, Hua-Min  
592 Zhong, Qiu-Lian Deng, Yan Long, Yi-Yu Yang, Si-Tang Gong and Zhen-Wen Zhou (2017)  
593 Antimicrobial resistance and prevalence of CvfB, SEK and SEQ genes among  
594 *Staphylococcus aureus* isolates from paediatric patients with bloodstream infections. *Exp*  
595 *Ther Med* 14: 5143-5148.
- 596 74. Voyich JM, Braughton KR, Sturdevant DE, Whitney AR, Said-Salim B, et al. (2005) Insights into  
597 mechanisms used by *Staphylococcus aureus* to avoid destruction by human neutrophils.  
598 *J Immunol* 175: 3907-3919.

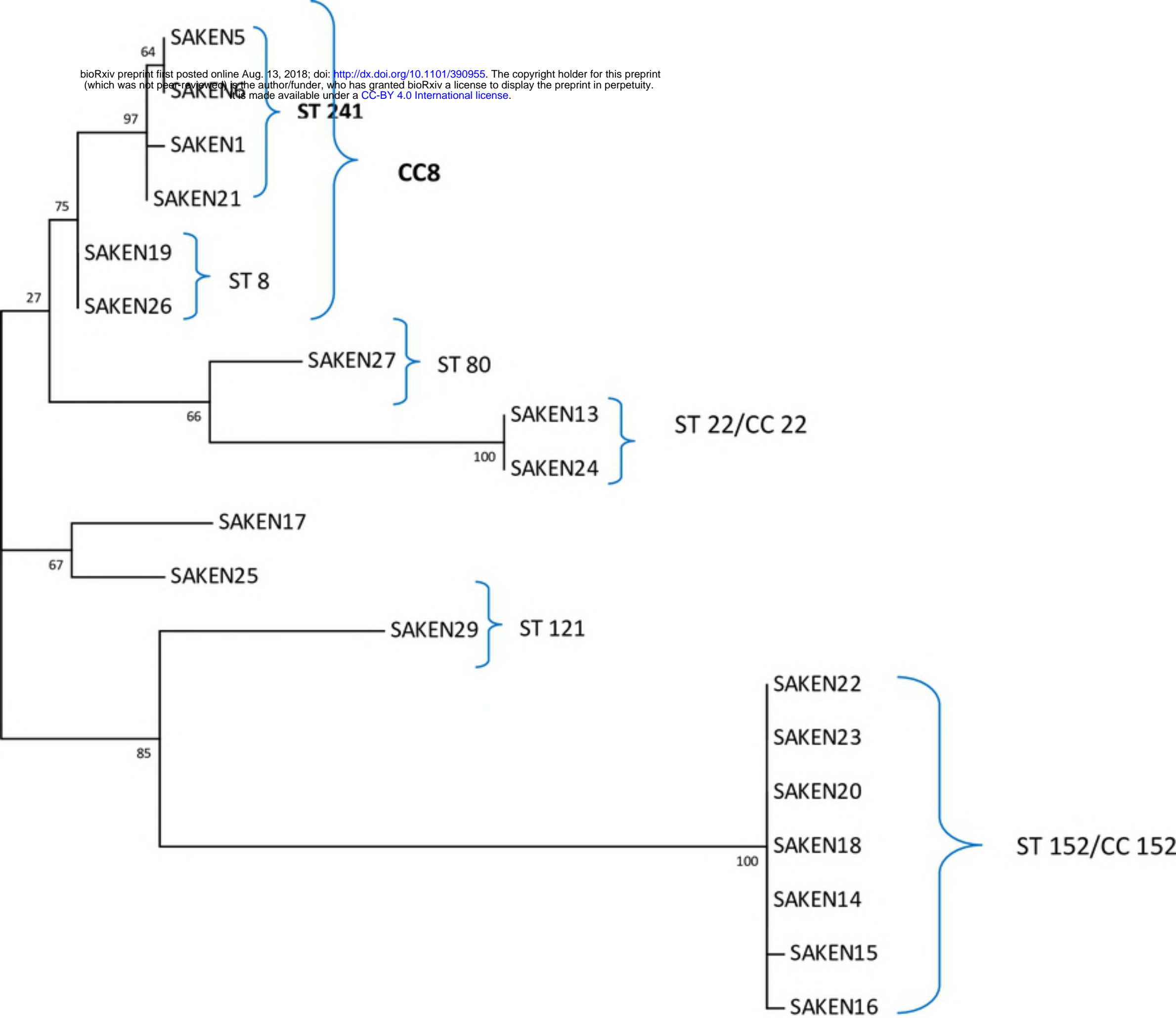
## 599 **SUPPLEMENTARY INFORMATION**

600 **S1 Table 1: List of reference genomes used in this study**

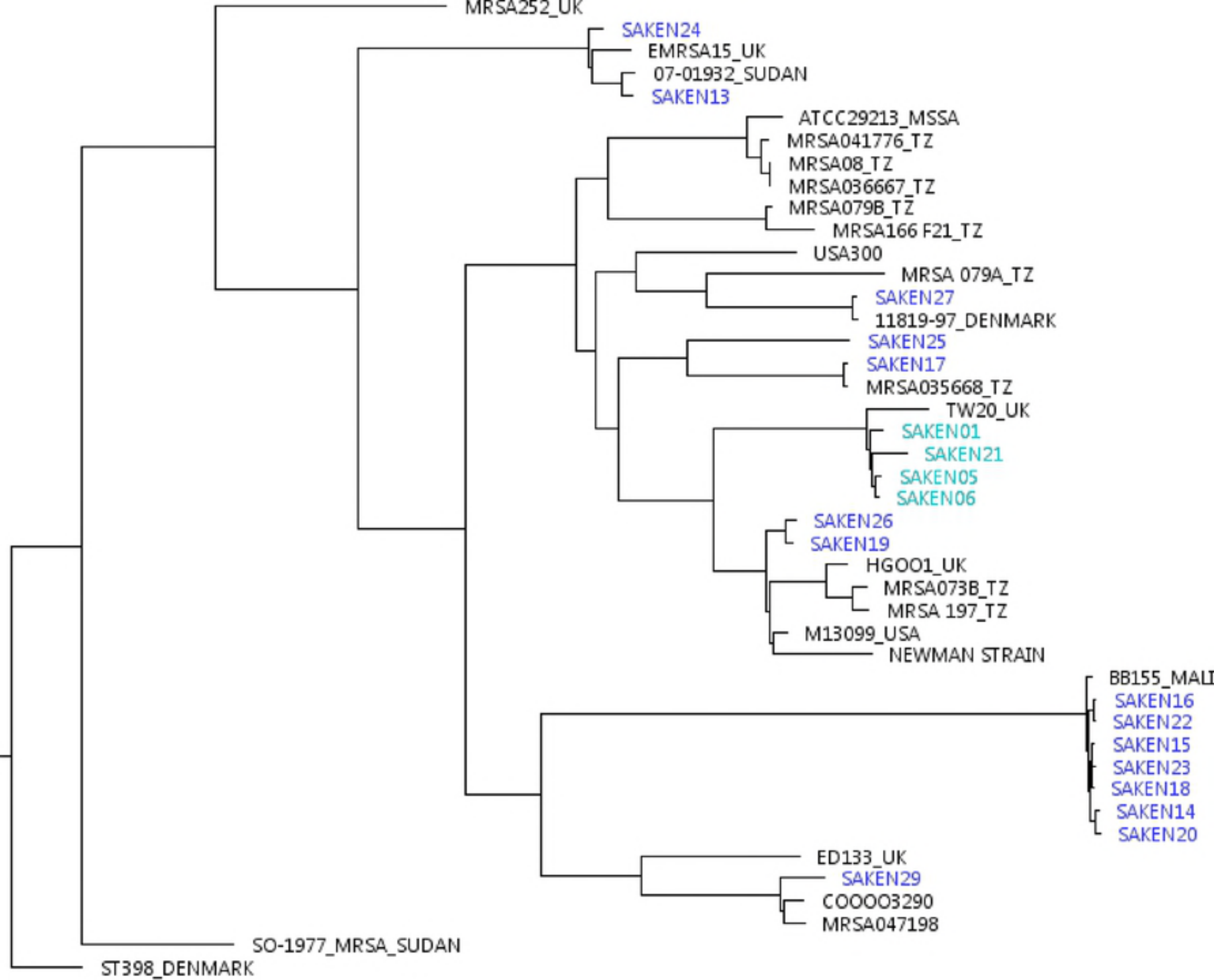
601 **S2 Table 2: Virulence gene profiles across *S. aureus* isolates**

602

603







0.003