bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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:** Phylogenomic study of *Staphylococcus aureus* and *Staphylococcus haemolyticus* clinical
2 isolates from Egypt

3 **Authors:** Cesar Montelongo[a], Carine R. Mores[a], Catherine Putonti[a,b,c], Alan J. Wolfe[a], Alaa
4 Abouelfetouh[d,e*]
5

6 **Affiliations:**
7 [a]Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University
8 Chicago, Maywood, IL 60153 USA
9 [b]Bioinformatics Program, Loyola University Chicago, Chicago, IL 60660 USA
10 [c]Department of Biology, Loyola University Chicago, Chicago, IL 60660 USA
11 [d]Department of Microbiology and Immunology, Faculty of Pharmacy, Alexandria University,
12 Alexandria, Egypt
13 [e]Department of Microbiology and Immunology, Faculty of Pharmacy, Alalamein International
14 University, Alamein, Egypt
15 **Author email addresses:**
16 Cesar Montelongo:
17 cmontelongoherna@luc.edu
18 Carine R. Mores:
19 carinermores@gmail.com
20 Catherine Putonti:
21 cputonti@luc.edu
22 Alan J. Wolfe:
23 Awolfe@luc.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

24 Alaa Abouelfetouh:

25 Alaa.abouelfetouh@pharmacy.alexu.edu.eg

26 aabouelfetouh@Aiu.edu.eg

27

28 **Corresponding Author:** Alaa Abouelfetouh; Alaa.abouelfetouh@pharmacy.alexu.edu.eg

29

30 **Keywords:** *Staphylococcus aureus*; *Staphylococcus haemolyticus*; Middle East; MLST

31

32 **Repositories:** Raw sequencing reads and assembled genomes can be found at BioProject

33 Accession number PRJNA648411 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA648411).

34

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

35 **Abstract**

36 Antibiotic resistant *Staphylococcus* infections are a global concern, with increasing cases of
37 resistant *Staphylococcus aureus* and *Staphylococcus haemolyticus* found circulating in the
38 Middle East. While extensive surveys have described the prevalence of resistant infections in
39 Europe, Asia, and North America, the population structure of resistant staphylococcal Middle
40 Eastern clinical isolates is poorly characterized. We performed whole genome sequencing of 56
41 *S. aureus* and 10 *S. haemolyticus* isolates from Alexandria Main University Hospital.
42 Supplemented with additional publicly available genomes from the region (34 *S. aureus* and 6 *S.*
43 *haemolyticus*), we present the largest genomic study of staphylococcal Middle Eastern isolates.
44 These genomes include 20 *S. aureus* multilocus sequence typing (MLST) types and 9 *S.*
45 *haemolyticus* MLSTs, including 3 and 1 new MLSTs, respectively. Phylogenomic analyses of
46 each species core genome largely mirrored MLSTs, irrespective of geographical origin. The
47 hospital-acquired *spa* t037/SCC*mec* III/MLST CC8 clone represented the largest clade,
48 comprising 22% of *S. aureus* isolates. Similar to other regional genome surveys of *S. aureus*, the
49 Middle Eastern isolates have an open pangenome, a strong indicator of gene exchange of
50 virulence factors and antibiotic resistance genes with other reservoirs. We recommend stricter
51 implementation of antibiotic stewardship and infection control plans in the region.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

52 **Impact Statement**

53 Staphylococci are under-studied despite their prevalence within the Middle East. Methicillin-
54 resistant *Staphylococcus aureus* (MRSA) is endemic to hospitals in this region, as are other
55 antibiotic-resistant strains of *S. aureus* and *S. haemolyticus*. To provide insight into the strains
56 currently in circulation within Egypt, we performed whole genome sequencing of 56 *S. aureus*
57 and 10 *S. haemolyticus* isolates from Alexandria Main University Hospital (AMUH). Through
58 analysis of these genomes, as well as other genomes of isolates from the Middle East, we were
59 able to produce a more complete picture of the current diversity than traditional molecular typing
60 strategies. Furthermore, the *S. haemolyticus* genome analyses provide the first insight into strains
61 found in Egypt. Our analysis of resistance and virulence mechanisms carried by these strains
62 provides invaluable insight into future plans of antibiotic stewardship and infection control
63 within the region.

64

65 **Data Summary**

66 Raw sequencing reads and assembled genomes can be found at BioProject Accession number
67 PRJNA648411 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA648411).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

68 **Introduction**

69 Staphylococci are a heterogenous group of commensal bacteria in humans with the
70 potential to cause infections [1]. Two staphylococcal species especially relevant to the clinical
71 setting are *Staphylococcus aureus* and *Staphylococcus haemolyticus*. *S. aureus* is arguably the
72 most clinically important staphylococcal species; the infections it can cause range from mild
73 erythema to serious life-threatening ailments, including septicemia, pneumonia, and endocarditis
74 [2]. A difficulty in treating and controlling *S. aureus* stems from its prevalence and increasing
75 resistance to clinically used antibiotics, resulting in it being one of the leading agents for
76 nosocomial and community-acquired infections [3,4]. *S. haemolyticus* is the second most common
77 staphylococcal species isolated in human blood cultures and a prominent reservoir for antibiotic
78 resistance genes, which can be shared with other Staphylococci, including *S. aureus* [5–7].

79 Epidemiological surveillance and profiling are key to managing Staphylococci [8,9].
80 Historically, profiling of Staphylococci has relied on complementary molecular typing strategies,
81 such as Multi Locus Sequence Typing (MLST), typing of hypervariable short repeats in Protein
82 A (*spa*), subtyping elements in the cassette chromosome mec (SCC*mec*), and presence of the
83 Panton-Valentine leucocidin (PVL) [10]. MLST consists of comparing the sequence of specific
84 housekeeping genes in bacteria; the strategy is effective at tracking a broad range of clones over
85 a global area, but prior to whole-genome sequencing (WGS) this method was slow and
86 expensive [10]. *spa* typing complements MLST by tracking the molecular evolution of *S. aureus*,
87 given the relevance of Protein A to the infectious process [9]. SCC*mec* permits profiling of
88 clinically relevant antibiotic resistances, including *mecA*, which results in methicillin-resistant *S.*
89 *aureus* (MRSA) [11,12]. PVL is an important cytotoxin in MRSA that is common in community-
90 acquired MRSA (CA-MRSA), but uncommon in hospital-associated MRSA (HA-MRSA) [11].
91 Traditionally, these profiling strategies have been a powerful means to type, trace, and manage

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

92 Staphylococcal infections, but technical limitations curtail the usefulness of molecular typing in
93 real time [9,10].
94 The increasing utility, speed, and inexpensiveness of WGS in the clinical setting is poised
95 to immensely benefit Staphylococci profiling [10,13–15]. WGS allows access to the entire
96 staphylococcal genome, including sequence data for typing MLST, *spa*, SCC*mec*, and PVL. In
97 addition, WGS allows us to study the phylogenomic lineage, core, and accessory genome of
98 isolates from an infectious outbreak or a geographical area. A key question is how WGS analysis
99 compares to traditional typing techniques. For example, there is evidence that phylogenomic data
100 does not always agree with standard typing methods; skepticism also exists that WGS can
101 reliably detect single nucleotide polymorphisms (SNPs) in sensitive genetic content [10]. In
102 contrast, studies have demonstrated that WGS can be used to type, discriminate, and cluster
103 staphylococcal isolates for the purpose of outbreak control [13–15]. WGS could be used to close the
104 gap in staphylococcal management in regions that have not been extensively monitored, such as
105 the Middle East and specifically Egypt.
106 The epidemiology of Staphylococci in non-European countries of the Mediterranean
107 region is under-studied [16]. Antibiotic resistance in *S. haemolyticus* has been identified in Middle
108 Eastern countries, such as Turkey [6] and Egypt [7]. There is evidence that MRSA is prevalent and
109 endemic to hospitals in this region, with a median MRSA prevalence of 38% in Algeria, Cyprus,
110 Egypt, Jordan, Lebanon, Malta, Morocco, Tunisia and Turkey [17]. Broadly speaking, PVL
111 prevalence is reported as low in some of these countries, indicating a predominance of HA-
112 MRSA [16,18,19]. Research into the lineage of Staphylococci in this region is urgent, as it would
113 give us both a present and future assessment of staphylococcal epidemiology.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

114 Generally, molecular typing and phylogeny data are limited from this region. Multiple
115 isolates in Palestine were typed as ST22 with a minority typed as ST80-MRSA-IV and PVL-
116 positive [20]. In Jordan, genotyping of *S. aureus* isolates revealed that the majority were ST80-
117 MRSA-IV [21]. In Lebanon, the primary lineage was PVL-positive ST80-MRSA-IV followed by
118 PVL-positive ST30-MSSA [17]. In Algeria, it was reported that ST80-MRSA-IV was present in
119 most neonates tested over an 18-month period, with a minority of these PVL-positive [22]. Finally,
120 for Egypt, it has been reported that the prevalent MLSTs are ST30, ST80, and a novel type,
121 ST1010; PVL prevalence has been estimated at 19% [23]. Enany *et al*. reported that the Egyptian
122 ST80 lineage was different from the globally prevalent ST80, primarily due to a unique *spa* type
123 and antimicrobial resistance [23].

124 Egypt presents a unique case-study for staphylococcal distribution in Arab countries [24].
125 Egypt's cultural and geographical placement may facilitate local Staphylococcal exposure to
126 international lineages, both from the Middle East and elsewhere . The accessibility of WGS
127 presents an opportunity to profile Staphylococci in Egypt and the rest of the Arab region in terms
128 of gene marker typing, core genome, and phylogenomics. Prior to this study, there were limited
129 genomic data of *S. aureus* and *S. haemolyticus* in this region.

130 Here, we report the phylogenetic and phylogenomic associations of 56 *S. aureus* and 10
131 *S. haemolyticus* isolates from Egypt and their relationship to 34 *S. aureus* and 6 *S. haemolyticus*
132 isolates from Egypt, Kuwait, Lebanon, Tunisia, Palestine, United Arab Emirates, Morocco, and
133 Sudan. WGS afforded insight into the lineage and genetic content of these two staphylococcal
134 species, including type information historically obtained using molecular methods. Both the
135 MLST and SCC*mec* type mirrored the core genome, indicating that WGS is a fast and accessible
136 option for Staphylococcal profiling. We identified multiple MLST and clonal complexes in

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

137 circulation in the region, including 3 new genotypes. Genome analysis indicated that *S. aureus* in
138 Egypt has an open pangenome that includes virulence genes in both the core and accessory
139 genomes. Surveillance and profiling of Staphylococci are key to infection control, and we have
140 shown that WGS can be a valuable asset, especially in regions where Staphylococci have not
141 been well studied, such as the Middle East.

142

143 **Results**

144 *S. aureus* and *S. haemolyticus* isolates were collected from patients presenting to the Medical
145 Microbiology Laboratory at AMUH between September and December 2015. Draft genomes for
146 66 of the clinical isolates were of high quality and were included in our analysis. These genomes
147 included 56 *S. aureus* and 10 *S. haemolyticus* isolates, assembled on average into 71 and 126
148 contigs, respectively. Additional publicly available *S. aureus* and *S. haemolyticus* strains were
149 identified and included in subsequent analyses: 34 *S. aureus* (from Egypt n=17; Kuwait n=5;
150 Lebanon n=4; Tunisia, Palestine and United Arab Emirates n=2 each; Morocco and Sudan n=1
151 each) and 6 *S. haemolyticus* (all from Egypt). **Supplementary Table S1** lists the available
152 metadata and presents the genome assembly statistics for *S. aureus* and *S. haemolyticus*. The *S.*
153 *aureus* genomes were, on average, larger than *S. haemolyticus* genomes: 2.8 Mbp and 2.5 Mbp,
154 respectively. Genome size and GC content were on par with other publicly available genomes.
155 Each draft genome was annotated using NCBI's PGAP, identifying an average of 2,839 and
156 2,495 coding sequences (CDS) for *S. aureus* and *S. haemolyticus*, respectively. The strains varied
157 in their number of rRNA operons and tRNAs.

158

159 ***Strain genotyping***

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

160 The genomes represent varied MLSTs. The 16 *S. haemolyticus* isolates examined here belonged
161 to nine MLSTs, including a new genotype ST-74 (strain 51) assigned as a result of this study,
162 and an isolate of unknown ST (strain 7A). ST-3 was the most common amongst the isolates
163 examined (n=4) (**Supplementary Table S2**). A total of 20 *S. aureus* MLSTs were identified,
164 including three novel types ST-5860 (strain 48), ST-5861 (strain 2705404), and ST-5862 (strain
165 2705410); all three of these strains came from prior studies and were isolated from Egypt,
166 Kuwait and Lebanon, respectively (**Supplementary Table S2**). Twelve different MLSTs were
167 identified among the Egyptian isolates; ST-239 was the most prevalent (n=24), followed by ST-1
168 (n=19) and then ST-80 (n=12). Two of the AMUH *S. aureus* isolates, strains AA32 and AA35,
169 could not be typed due to incomplete sequences.

170 *S. aureus* isolates could be categorized into seven clonal complexes (CC)
171 (**Supplementary Table S2**), the largest being CC8 (n=26), consisting mainly of Egyptian
172 isolates and one Moroccan isolate (strain 12480433). 12 strains were identified as ST-80, which
173 does not belong to a clonal complex. 20 different *spa* types were identified in addition to 7
174 isolates that could not be typed. The predominant *spa* type was t037 (n=33), with all but one
175 belonging to CC8; 30 of these 33 isolates belonged to SCC*mec* III. The next most frequent *spa*
176 type was t127 (n=19), all belonging to CC1. **Table 1** summarizes these results.

177

178 ***Core and pangenomes of* S. haemolyticus *and* S. aureus *strains***

179 To investigate the core genome and pangenome of *S. haemolyticus*, we added the 6 publicly
180 available *S. haemolyticus* genomes from Egypt to our 10 *S. haemolyticus* genomes
181 (**Supplementary Table S1**). The pangenome for these strains included 3,541 genes
182 (**Supplementary Fig. S1**), with 1,834 single copy number genes in the core genome. Included

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

183 within these core genes are the virulence factors autolysin (*atl*), elastin binding protein (*ebp*),
184 thermonuclease (*nuc*), and cytolysin (*cylR2*).

185 In addition to our 56 *S. aureus* genomes, our core and pangenome analysis included 34 *S.*
186 *aureus* draft genome assemblies from the Arab region (**Supplementary Table S1**). The
187 pangenome of these 90 isolates contained 4,283 genes (**Fig. 1, panel A**), the core genome
188 included 1,501 single copy number genes, and the accessory genome contained 2,178 genes.
189 These analyses show that the Arab isolates have an open pangenome. The functionality of the
190 genes within the *S. aureus* core genome was determined according to their COG categories (**Fig.**
191 **1, panel B**). The core genome was further examined for virulence factors, finding the same gene
192 related to autolysin (*atl*) that was found in the *S. haemolyticus* core. We also identified genes
193 associated with intercellular adhesin, cysteine protease, thermonuclease, capsule, and the Type
194 VII secretion system (**Table 2**).

195

196 In addition to the virulence factors found within the core genome, we identified virulence
197 factors and antibiotic resistance genes within the accessory genome (**Supplementary Tables S3**
198 **and S4**). The isolates were screened for the presence of *lukF/S*-PV, which encodes PVL.
199 Isolates positive for PVL were mainly (77%) *mecA* positive, present in CC1, ST-80, CC30 and
200 CC8, and obtained from Egypt, Kuwait, Tunisia, Lebanon and Morocco. Importantly, isolates
201 obtained from CA infections belonged to CC1, ST-80, CC5, CC97 and CC8, making PVL
202 presence a good predictor for the ability of the isolate to cause CA infections.

203

204

205 ***Phylogenomic study of* S. haemolyticus *and* S. aureus *strains from the Arab region***

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

206 The core genes were used to derive phylogenies for each species. The *S. haemolyticus* isolates
207 were all from Egypt and clustered into two clades (**Fig. 2**). As the tree shows, variation between
208 the core genomes of these isolates was minor. Furthermore, the clade structure of the genomes
209 corresponded with MLST, indicated in the bar of **Fig. 2**. The MLST tree for these genomes is
210 shown in **Supplementary Fig. S2**.
211
212 *S. aureus* isolates came from all over the region, and clustered into six clades, with Egyptian
213 isolates represented in all clades (**Fig. 3 and 4**). Clade 1 isolates belonged to ST-1 and were from
214 Egypt and the UAE, clade 2 contained the majority of the Arab isolates including some from
215 Egypt. The predominating clone seen among 46.7% of the isolates within clade 2 was *spa*
216 t044/SCC*mec* IV/ST-80, which shows some degree of shared content between these isolates.
217 Clade 3 isolates were solely from Egypt and belonged mainly to ST-15 and ST-5. Clade 4
218 comprised isolates from Egypt, Sudan and Palestine, with the majority belonging to ST-22 and
219 ST-361. Clade 5 contained isolates from Egypt, belonging mainly to ST-97. The remaining
220 isolates were in clade 6, of ST-239 and from Egypt, with the exception of one Moroccan isolate.
221 This clade represents a *spa* t037/SCC*mec* III/MLST CC8 clone. The phylogenetic tree derived
222 from the core genome sequences corresponded with the tree derived from the MLST marker
223 genes (**Fig. 3** and **Supplementary Fig. S2**). 14 isolates lacked *mecA* (**Fig. 4**, pale green star) and
224 occurred predominantly in CC1 (n=5), CC15 (n=3) and CC30, CC8 and ST-80, with one isolate
225 in each; in addition, three isolates belonged to ST-361 (n=2) or ST-5860 (n=1). 13 of these *mecA*
226 negative isolates were from Egypt.
227
228

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

229

230 **Discussion**

231 *S. aureus* is a major human pathogen in hospital and community settings, with the
232 infection rate of MRSA increasing on a global scale at varying rates [25,26]. While extensive
233 surveys have provided insight into the prevalence of such resistant infections in Europe [27], Asia
234 [28,29], and North America [30,31], limited data is available for the Arab region [32]. Furthermore,
235 antibiotic resistant *S. haemolyticus* strains have been identified worldwide [5], including Turkey [6]
236 and Egypt [7]. Prior to the study initiated here, there were limited genomic data for these two
237 *Staphylococcus* species from the Middle East. The addition of 56 *S. aureus* and 10 *S.*
238 *haemolyticus* genomes enabled our investigation of strain diversity within the region. With the
239 majority (or all, in the case of *S. haemolyticus*) of genomes representing isolates from Egypt, we
240 could investigate members of these species currently in circulation. We found that several
241 different MLSTs and clonal complexes are in circulation within Egypt and more broadly within
242 the region. 20 *S. aureus* MLSTs were identified in the region, including 3 new genotypes
243 identified here, and 12 of these are in circulation within Egypt. Analysis of the *S. haemolyticus*
244 genomes found 9 MLSTs in circulation within Egypt, including one new genotype.

245 The core genome for the *S. aureus* strains is slightly larger than that previously calculated
246 for the species [33,34]. This is expected, however, as our analysis is restricted to fewer genomes
247 from a single region. Both our *S. haemolyticus* and *S. aureus* core genomes include the gene *atl*,
248 which is essential for biofilm formation [35]. Furthermore, the *ica* gene cluster, also associated
249 with biofilm formation [36], as well as its regulator *icaR* [37], are included in the core genome of the
250 *S. aureus* strains examined here. The presence of *atl* and the *ica* gene clusters signifies the
251 biofilm potential of the isolates. This potential is relevant because 80% of human microbial

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

252 infections are complicated by biofilm formation, such as in wounds, IV catheters, sutures and
253 implants (see reviews [38,39]). Moreover, the biofilm capacity to evade the host's immune defenses,
254 the inability of most antibiotic treatment regimens to eradicate existing biofilms and the fact that
255 biofilms serve as a good medium for exchange of genetic material (e.g. plasmids) between cells
256 make biofilm formation a major health concern in the clinical setting ([40]; see reviews [41,42]).

257 The Arab *S. aureus* genomes have an open pangenome, evidence of gene exchange
258 between these isolates and other reservoirs. *S. aureus* is naturally competent [43], and horizontal
259 gene transfer (HGT) between strains, coagulase-negative *Staphylococcus* (CoNS) strains, and
260 other species is well documented (see review [44]). Recently, HGT was shown to be a driver of
261 persistent *S. aureus* infections within patients [45]. Genes within the accessory genome included
262 virulence factors and antibiotic resistance genes (**Supplementary Tables S3 and S4**). Prior
263 comparative genomic studies for this species similarly found an open pangenome and resistance
264 genes within the accessory genome [33].

265 Phylogenomic analyses of the core genome largely mirrored MLST types (**Fig. 3**). This
266 was irrespective of geographical origin. Interestingly, strains of the same SCC*mec* type had a
267 more similar core genome sequence (**Fig. 4**). In a prior phylogenetic study, John and co-workers
268 found that 16S rRNA gene sequence similarity did not correspond with SCC*mec* type, leading
269 them to conclude that horizontal gene transfer plays a role in resistance gene acquisition [33].
270 However, only two SCC*mec* types were identified for the samples examined here. Recently,
271 Soliman *et al*. published a study characterizing the genomes of 18 MRSA isolates from a tertiary
272 care hospital in Cairo, Egypt; their isolates were primarily SCC*mec* types V (n=9) and VI (n=2),
273 not observed within our larger collection [46]. Similarly, SCC*mec* type V and IV have been most
274 frequently observed in other *S. aureus* studies within the region [47–49]. Rather, our study found

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

275 that SCC*mec* type III and IV were equally prevalent within the region. SCC*mec* type III
276 predominated among HA-MRSA infections, suggesting that, in contrast to these prior studies,
277 our isolates indicate that HA-infections have a higher incidence among the patients tested here.
278 AMUH, where our isolates were collected, is the largest tertiary hospital and main referral center
279 in the Northern sector of Egypt; thus, patients with more severe infections would be more likely
280 to be treated at AMUH than at any other hospital in the region. Prior studies have found SCC*mec*
281 type III to be the predominant type in Asian countries [28]. Besides, the SCC*mec* type III/MLST
282 ST-239 is the oldest pandemic strain of MRSA [50], which might explain its prevalence among the
283 current collection of isolates.

284 The *S. haemolyticus* and *S. aureus* genomes examined here provide insight into the
285 diversity of strains currently in circulation in Egypt, particularly with respect to their encoded
286 virulence factors and antibiotic resistance genes. WGS analysis enabled a more complete picture
287 of this diversity than molecular typing strategies. The *S. haemolyticus* genomes provide the first
288 insight into strains found in Egypt. Identifying the main genotypes, as well as the resistance and
289 virulence mechanisms among the resistant isolates in the region, can drive antibiotic stewardship
290 and infection control plans.

291

292 **Methods**

293 Bacterial isolates

294 A total of 89 *S. aureus* and 14 *S. haemolyticus* consecutive non-duplicate isolates were collected
295 from the Medical Microbiology Laboratory at Alexandria Main University Hospital (AMUH)
296 between September and December 2015. These isolates were obtained from various clinical
297 specimens, including pus, blood, sputum, urine, tissue, aspirate and broncho-alveolar lavage

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

298 (BAL). The identity of the isolates was determined using conventional methods, such as Gram
299 staining, growth on and fermentation of mannitol salt agar, growth on DNase agar and slide
300 coagulase testing using Dryspot Staphytect Plus (Oxoid Ltd, England), and confirmed using
301 Matrix Assisted Laser Desorption Ionization - Time of Flight Mass Spectrometry (MALDI-TOF
302 MS) (Bruker Daltonik, USA). The isolates were further classified as either hospital-acquired or
303 community-acquired infections based on a 48 h window between the dates of patient admission
304 and isolate collection [51].
305
306 DNA extraction
307 Colonies grown on tryptone soya agar (TSA) plates were harvested and washed in 1 ml
308 phosphate buffer saline (PBS) and resuspended in 0.5 ml SET (75mM NaCl, 25mM EDTA,
309 20mM Tris, pH 7.5), to which 50ul of fresh 20 mg/ml lysozyme in PBS and 30ul Mutanolysin
310 were added; the mixture was incubated at 37ºC for 60 min. The cells were then treated with 60ul
311 10% sodium dodecyl sulphate and 20ul proteinase K and incubated at 55ºC for two hours with
312 gentle inversion. The suspension was mixed gently with 210ul of 6M NaCl, and 700ul
313 phenol:chloroform were added, followed by incubation at room temperature for 30-60 minutes,
314 using a rotating wheel for gentle mixing. The suspension was then centrifuged at maximum
315 speed for 10 min and the aqueous phase was transferred to a new microfuge tube and mixed
316 gently with an equal volume of isopropanol. The tubes were centrifuged to produce a DNA pellet
317 that was washed with 70% ethanol, which was left to evaporate overnight. The pellets were
318 resuspended in 50ul ddH_2O and stored at -20ºC till further processing.
319
320 Genome Sequencing and Genome Assembly

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

321 The Illumina Nextera kit was used for whole genome library preparation. Each isolate was
322 sequenced using the Illumina MiSeq System, producing paired-end 2x250 bp reads. Quality
323 control and de-multiplexing of sequence data was done with onboard MiSeq Control software
324 and MiSeq Reporter v3.1. Raw reads were trimmed using Sickle v1.33
325 (https://github.com/najoshi/sickle) and assembled using SPAdes v3.13.0 [52] with the "only-
326 assembler" option for k□=□55, 77, 99, and 127. Genome coverage was calculated using BBMap
327 v38.47 (https://sourceforge.net/projects/bbmap/). Contigs shorter than 500 bp were pruned using
328 bioawk (https://github.com/lh3/bioawk). Genome assemblies were annotated using PATRIC
329 v3.3.18 [53]. Genomes were deposited in NCBI's Assembly database, along with raw sequence
330 data in SRA under BioProject PRJNA648411. Deposited genomes were annotated using the
331 NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.0 [54]. Unless previously noted,
332 default parameters were used for each software tool. To complement our analysis of the genomes
333 from AMUH, raw sequence data for 41 *S. aureus* and 10 *S. haemolyticus* strains were retrieved
334 from NCBI. These records were identified by searching SRA (as of January 2020) for strains
335 isolated in the Arab region. These raw reads were processed as indicated above. High-quality
336 assemblies were included in subsequent analyses.
337
338 Bioinformatic Analysis
339 Multilocus sequence typing (MLST) was determined using the MLST v2.0.4 web server
340 available through the Center for Genomic Epidemiology [55]. MLST allele sequence and profile
341 data were obtained from PubMLST v2.0.0 [56]. *spa* typing was performed using the online tool
342 SpaTyper v1.0 available through the Center for Genomic Epidemiology [57]. SCC*mec* typing was
343 performed using SCCmecFinder v1.2 online tool available through the Center for Genomic

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

344 Epidemiology (https://cge.cbs.dtu.dk/services/SCCmecFinder/) [58,59]. Resistance and virulence
345 genes were identified using PATRIC v3.6.5 [60] and VFAnalyzer [61].
346
347 Phylogenomic and Phylogenetic Analysis
348 The core and pangenomes were generated using anvi'o v5.1. The following scripts were used to
349 calculate the pangenome: anvi-gen-genomes-storage, anvi-pan-genome and anvi-display-pan,
350 and the following script was used to calculate the core genome: anvi-get-sequences-for-gene-
351 clusters [62,63]. Functional groups for the core genome were determined by querying core genome
352 amino acid sequences against the COG database [64] through anvi'o. The core genes were
353 concatenated for each genome and then aligned using MAFFT v7.388 [65]. The tree was built
354 using the FastTree v2 [66] plugin in Geneious Prime v2019.2 (Biomatters Ltd., Auckland, New
355 Zealand). MLST ST sequences were downloaded from PubMLST v2.0.0 [56], aligned in Geneious
356 Prime v2019.2 and the trees were built using the FastTree v2 [66] plugin in Geneious Prime
357 v2019.2 . iTOL v5.6.1 [67] was used to annotate and visualize all trees.
358
359
360 **Author Statements**
361
362 Acknowledgments
363 We acknowledge Roberto Limeira and Loyola Genomics Facility for performing the whole
364 genome sequencing of the isolates. We also acknowledge funding from NIH (R01 DK104718
365 awarded to AJW), NSF (1661357 awarded to CP), USAID (GSP-T85 awarded to AA) and DFG
366 (ZI 665/3-1 awarded to AA). The funders did not play a part in the design or conduct of the study

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

367 Authors and Contributors

368 CM: Formal Analysis; Writing – Original Draft Preparation; Writing – Review and Editing

369 CRM: Formal Analysis; Writing – Original Draft Preparation; Writing – Review and Editing

370 CP: Formal Analysis; Writing – Original Draft Preparation; Visualization; Writing – Review and

371 Editing

372 AJW: Formal Analysis; Conceptualisation; Writing – Review and Editing

373 AA: Conceptualisation; Formal Analysis; Writing – Original Draft Preparation; Writing –

374 Review and Editing

375 All authors reviewed the manuscript.

376 Competing Interests

377 AJW is a member of the Advisory Board of Urobiome Therapeutics. The remaining authors

378 report no disclosures.

379 Data availability

380 Raw sequencing reads and assembled genomes can be found at BioProject Accession number

381 PRJNA648411 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA648411)

382 Ethics declarations:

383 Ethics approval and consent to participate

384 Not applicable

385

386 Consent for publication

387 Not applicable

388

389 Figure legends

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

390 **Figure 1.** Genome analysis of 90 Arab *S. aureus* strains. **(a)** The pangenome. Each ring
391 corresponds to a single genome. Each radial extension in the ring corresponds to the presence
392 (black) or absence (light gray) of a given gene cluster (homologous gene). The bar charts list the
393 number of genes identified in the given genome (top) and the number of singleton genes or genes
394 that are unique to the given genome (bottom). The pangenome of these 90 isolates contained
395 4,283 genes, the core genome included 1,501 single copy number genes, and the accessory
396 genome contained 2,178 genes. **(b)** Functionality of genes contained within the core genome.
397 The same autolysin gene (*atl*) found in the core genome of *S. haemolyticus* was found in *S.*
398 *aureus*.
399
400 **Figure 2.** Phylogeny based upon the core genes for the *S. haemolyticus* isolates. All *S.*
401 *haemolyticus* isolates were from Egypt and clustered into two clades corresponding with MLST.
402
403 **Figure 3.** *S. aureus* core genome phylogeny colored by geographical origin of isolation (strain
404 name color) and MLST (right bar). *S. aureus* isolates were from different parts of the region, and
405 clustered into six clades, each containing Egyptian isolates. Clade 1 isolates belonged to ST-1
406 and were from Egypt and the UAE, clade 2 contained the majority of the Arab isolates, with *spa*
407 t044/SCC*mec* IV/ST-80 as the predominating clone. Clade 3 isolates were solely from Egypt and
408 belonged mainly to ST-15 and ST-5. Clade 4 comprised isolates from Egypt, Sudan and
409 Palestine, with the majority belonging to ST-22 and ST-361. Clade 5 contained isolates from
410 Egypt and belonged mainly to ST-97. The remaining isolates were in clade 6, of ST-239 and
411 from Egypt and Morocco (n=1). This clade represents a *spa* t037/SCC*mec* III/MLST CC8 clone.
412

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

413 **Figure 4.** Phylogenetic tree based on core genome annotated by geographic origin, MLST CC,
414 main *spa* and SCC*mec* types. 14 isolates, mostly from Egypt, lacked *mecA* and occurred
415 predominantly in CC1 (n=5), CC15 (n=3) and CC30, CC8 and ST-80 (one isolate in each), ST-
416 361 (n=2) or ST-5860 (n=1).

417

418

419 **Table 1:** MLST clonal complexes, *spa* types, and SCC*mec* types among the *S. aureus* isolates.

MLST CC	**Strain Name**	**Geographical Origin**	***spa* type**	**SCC*mec* type**
CC1	3 (A), 3 (B), 23, 6 (B), 43, AA51, AA67, AA77	Egypt	t127	N/D
	R181, R180, AA1, AA78	UAE, Egypt	t127	N/D
	6 (A), AA69	Egypt	t127	N/D
	AA59, AA65, AA68	Egypt	t127	predicted as MSSA
	AA103, AA87	Egypt	t127	predicted as MSSA
CC15	15, 16	Egypt	t094	predicted as MSSA
	17	Egypt	unk	predicted as MSSA
CC22	41	Egypt	t13828	IV
	Gaza_MRSA_B62	Palestine	t223	IV
	AA18	Egypt	t223	IV
	AA5	Egypt	t3243	IV
	Gaza_MRSA_B04	Palestine	t790	IV
	40	Egypt	unk	IV
CC30	AA41	Egypt	t037	predicted as MSSA
	19	Egypt	t1504	N/D

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

CC5	AA30	Egypt	t304	IV
	14, AA76, AA80	Egypt	t688	N/D
	AA70	Egypt	t688	VI(4B)
CC8	12480433	Morocco	t008	IV
	LHI_Sa_30	Egypt	t008	N/D
	46	Egypt	t030	III
	50, AA101, AA13, AA14, AA22, AA23, AA27, AA31, AA33, AA46, AA52, AA55, AA57, AA60, AA61, AA62, AA63, AA64, AA91, AA92	Egypt	t037	III
	AA79	Egypt	t037	N/D
	AA93	Egypt	t037	predicted as MSSA
	AA29	Egypt	unk	III
CC97	AA36	Egypt	t267	N/D
	AA39, AA6	Egypt	t267	IV
	AA104	Egypt	t267	N/D
	AA8	Egypt	unk	IV
ST-80	2705432, 2705403, 2705405, 2705407, 2705409, 2705412	Tunisia, Kuwait, Lebanon	t044	IV
	AA45	Egypt	t044	IV
	2705431	Tunisia	t044	predicted as MSSA
	2705411	Lebanon	t131	IV
	AA2	Egypt	t416	IV
	AA3, AA4	Egypt	t416	N/D

420

421 **Table 2.** Virulence factors included in the *S. aureus* core genome.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

VFclass	**Virulence factors**	**Related genes**
Adherence	Autolysin	*atl*
	Intercellular adhesin	*icaA*
		icaD
		icaR
Enzyme	Cysteine protease	*sspC*
	Thermonuclease	*nuc*
Immune evasion	Capsule	*cap5A*
		cap8B
		cap5M
		cap8N
		capO
Secretion system	Type VII secretion system	*esaB*
		essA
		essB
		esxA

422

423

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

424 **References**

425 1. Becker, K., Heilmann, C. & Peters, G. Coagulase-negative staphylococci. *Clinical*
426 *Microbiology Reviews* **27**, 870–926 (2014).
427 2. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G.
428 Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations,
429 and management. *Clinical Microbiology Reviews* **28**, 603–661 (2015).
430 3. Stefani, S. *et al.* Meticillin-resistant Staphylococcus aureus (MRSA): Global epidemiology
431 and harmonisation of typing methods. *International Journal of Antimicrobial Agents* vol. 39
432 273–282 (2012).
433 4. Chambers, H. F. & DeLeo, F. R. Waves of resistance: Staphylococcus aureus in the
434 antibiotic era. *Nature Reviews Microbiology* vol. 7 629–641 (2009).
435 5. Froggatt, J. W., Johnston, J. L., Galetto, D. W. & Archer, G. L. Antimicrobial resistance in
436 nosocomial isolates of Staphylococcus haemolyticus. *Antimicrobial Agents and*
437 *Chemotherapy* **33**, 460–466 (1989).
438 6. Morrissey, I., Leakey, A. & Northwood, J. B. In vitro activity of ceftaroline and comparator
439 antimicrobials against European and Middle East isolates from complicated skin and skin-
440 structure infections collected in 2008-2009. *International Journal of Antimicrobial Agents*
441 **40**, 227–234 (2012).
442 7. Maarouf, L., Omar, H., El-Nakeeb, M. & Abouelfetouh, A. Prevalence and mechanisms of
443 linezolid resistance among staphylococcal clinical isolates from Egypt. *European Journal of*
444 *Clinical Microbiology and Infectious Diseases* (2020) doi:10.1007/s10096-020-04045-w.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

445 8. Steinig, E. J. *et al.* Evolution and global transmission of a multidrug-resistant, community-
446 associated methicillin-resistant staphylococcus aureus lineage from the Indian subcontinent.
447 *mBio* **10**, (2019).
448 9. Pobiega, M., Wójkowska-Mach, J. & Heczko, P. B. Typing of Staphylococcus aureus in
449 order to determine the spread of drug resistant strains inside and outside hospital
450 environment. *Przeglad epidemiologiczny* **67**, 435-438,539-542 (2013).
451 10. Lindsay, J. A. Evolution of Staphylococcus aureus and MRSA during outbreaks. *Infection,*
452 *Genetics and Evolution* **21**, 548–553 (2014).
453 11. Funaki, T. *et al.* SCCmec typing of PVL-positive community-acquired Staphylococcus
454 aureus (CA-MRSA) at a Japanese hospital. *Heliyon* **5**, e01415 (2019).
455 12. Sabri, I., Adwan, K., Essawi, T. A. & Farraj, M. A. Molecular characterization of
456 methicillin-resistant Staphylococcus aureus isolates in three different Arab world countries .
457 *European Journal of Microbiology and Immunology* **3**, 183–187 (2013).
458 13. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-
459 resistant Staphylococcus aureus: A descriptive study. *The Lancet Infectious Diseases* **13**,
460 130–136 (2013).
461 14. Eyre, D. W. *et al.* A pilot study of rapid benchtop sequencing of Staphylococcus aureus and
462 Clostridium difficile for outbreak detection and surveillance. *BMJ Open* **2**, e001124 (2012).
463 15. Köser, C. U. *et al.* Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA
464 Outbreak. *New England Journal of Medicine* **366**, 2267–2275 (2012).
465 16. Tokajian, S. New epidemiology of Staphylococcus aureus infections in the Middle East.
466 *Clinical Microbiology and Infection* vol. 20 624–628 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

467 17. Borg, M. A. *et al.* Prevalence of methicillin-resistant Staphylococcus aureus (MRSA) in
468 invasive isolates from southern and eastern Mediterranean countries on behalf of the ARMed
469 Project members and collaborators. doi:10.1093/jac/dkm365.
470 18. Bhatta, D. R. *et al.* Association of Panton Valentine Leukocidin (PVL) genes with
471 methicillin resistant Staphylococcus aureus (MRSA) in Western Nepal: A matter of concern
472 for community infections (a hospital based prospective study). *BMC Infectious Diseases* **16**,
473 (2016).
474 19. Falagas, M. E., Karageorgopoulos, D. E., Leptidis, J. & Korbila, I. P. MRSA in Africa:
475 Filling the Global Map of Antimicrobial Resistance. *PLoS ONE* **8**, (2013).
476 20. Biber, A. *et al.* A typical hospital-acquired methicillin-resistant staphylococcus aureus clone
477 is widespread in the community in the Gaza strip. *PLoS ONE* **7**, (2012).
478 21. Khalil, W., Hashwa, F., Shihabi, A. & Tokajian, S. Methicillin-resistant Staphylococcus
479 aureus ST80-IV clone in children from Jordan. *Diagnostic Microbiology and Infectious*
480 *Disease* **73**, 228–230 (2012).
481 22. Djoudi, F. *et al. Panton-Valentine leukocidin positive sequence type 80 methicillin-resistant*
482 *Staphylococcus aureus carrying a staphylococcal cassette chromosome mec type IVc is*
483 *dominant in neonates and children in an Algiers hospital*. *NEW MICROBIOLOGICA* vol. 36
484 (2013).
485 23. Enany, S., Yaoita, E., Yoshida, Y., Enany, M. & Yamamoto, T. Molecular characterization
486 of Panton-Valentine leukocidin-positive community-acquired methicillin-resistant
487 Staphylococcus aureus isolates in Egypt. *Microbiological Research* **165**, 152–162 (2010).
488 24. Abouelfetouh, A. The Status of Methicillin Resistance Among Egyptian Staphylococcus
489 aureus Isolates: An Overview. *Infectious disorders drug targets* **17**, 67–69 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

490 25. Deurenberg, R. H. *et al.* The molecular evolution of methicillin-resistant Staphylococcus
491 aureus. *Clin Microbiol Infect* **13**, 222–235 (2007).
492 26. Guo, Y., Song, G., Sun, M., Wang, J. & Wang, Y. Prevalence and Therapies of Antibiotic-
493 Resistance in Staphylococcus aureus. *Frontiers in Cellular and Infection Microbiology* **10**,
494 (2020).
495 27. Köck, R. *et al.* Methicillin-resistant Staphylococcus aureus (MRSA): burden of disease and
496 control challenges in Europe. *Euro Surveill* **15**, 19688 (2010).
497 28. Chongtrakool, P. *et al.* Staphylococcal cassette chromosome mec (SCCmec) typing of
498 methicillin-resistant Staphylococcus aureus strains isolated in 11 Asian countries: a proposal
499 for a new nomenclature for SCCmec elements. *Antimicrob Agents Chemother* **50**, 1001–
500 1012 (2006).
501 29. Nickerson, E. K., West, T. E., Day, N. P. & Peacock, S. J. Staphylococcus aureus disease
502 and drug resistance in resource-limited countries in south and east Asia. *Lancet Infect Dis* **9**,
503 130–135 (2009).
504 30. Kallen, A. J. *et al.* Health care-associated invasive MRSA infections, 2005-2008. *JAMA* **304**,
505 641–648 (2010).
506 31. Laupland, K. B. *et al.* The changing epidemiology of Staphylococcus aureus bloodstream
507 infection: a multinational population-based surveillance study. *Clin Microbiol Infect* **19**,
508 465–471 (2013).
509 32. Yezli, S., Shibl, A. M., Livermore, D. M. & Memish, Z. A. Antimicrobial resistance among
510 Gram-positive pathogens in Saudi Arabia. *J Chemother* **24**, 125–136 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

511 33. John, J., George, S., Nori, S. R. C. & Nelson-Sathi, S. Phylogenomic Analysis Reveals the
512 Evolutionary Route of Resistant Genes in Staphylococcus aureus. *Genome Biol Evol* **11**,
513 2917–2926 (2019).
514 34. Nguyen, M., Olson, R., Shukla, M., VanOeffelen, M. & Davis, J. J. Predicting antimicrobial
515 resistance using conserved genes. *PLoS Comput Biol* **16**, e1008319 (2020).
516 35. Biswas, R. *et al.* Activity of the major staphylococcal autolysin Atl. *FEMS Microbiol Lett*
517 **259**, 260–268 (2006).
518 36. O'Gara, J. P. ica and beyond: biofilm mechanisms and regulation in Staphylococcus
519 epidermidis and Staphylococcus aureus. *FEMS Microbiol Lett* **270**, 179–188 (2007).
520 37. Jefferson, K. K., Pier, D. B., Goldmann, D. A. & Pier, G. B. The teicoplanin-associated locus
521 regulator (TcaR) and the intercellular adhesin locus regulator (IcaR) are transcriptional
522 inhibitors of the ica locus in Staphylococcus aureus. *J Bacteriol* **186**, 2449–2456 (2004).
523 38. Joo, H.-S. & Otto, M. Molecular basis of in vivo biofilm formation by bacterial pathogens.
524 *Chem Biol* **19**, 1503–1513 (2012).
525 39. Khatoon, Z., McTiernan, C. D., Suuronen, E. J., Mah, T.-F. & Alarcon, E. I. Bacterial
526 biofilm formation on implantable devices and approaches to its treatment and prevention.
527 *Heliyon* **4**, e01067 (2018).
528 40. Pereira-Ribeiro, P. M. *et al.* Influence of antibiotics on biofilm formation by different clones
529 of nosocomial Staphylococcus haemolyticus. *Future Microbiol* **14**, 789–799 (2019).
530 41. Donlan, R. M. Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**, 881–890 (2002).
531 42. Archer, N. K. *et al.* Staphylococcus aureus biofilms: properties, regulation, and roles in
532 human disease. *Virulence* **2**, 445–459 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

533 43. Morikawa, K. *et al.* Expression of a cryptic secondary sigma factor gene unveils natural
534 competence for DNA transformation in Staphylococcus aureus. *PLoS Pathog* **8**, e1003003
535 (2012).
536 44. Lindsay, J. A. Genomic variation and evolution of Staphylococcus aureus. *Int J Med*
537 *Microbiol* **300**, 98–103 (2010).
538 45. Langhanki, L. *et al.* In vivo competition and horizontal gene transfer among distinct
539 Staphylococcus aureus lineages as major drivers for adaptational changes during long-term
540 persistence in humans. *BMC Microbiol* **18**, 152 (2018).
541 46. Soliman, M. S. *et al.* Genomic Characterization of Methicillin-Resistant Staphylococcus
542 aureus (MRSA) by High-Throughput Sequencing in a Tertiary Care Hospital. *Genes (Basel)*
543 **11**, (2020).
544 47. Abou Shady, H. M., Bakr, A. E. A., Hashad, M. E. & Alzohairy, M. A. Staphylococcus
545 aureus nasal carriage among outpatients attending primary health care centers: a comparative
546 study of two cities in Saudi Arabia and Egypt. *Braz J Infect Dis* **19**, 68–76 (2015).
547 48. Alkharsah, K. R. *et al.* Comparative and molecular analysis of MRSA isolates from infection
548 sites and carrier colonization sites. *Ann Clin Microbiol Antimicrob* **17**, 7 (2018).
549 49. Hadyeh, E., Azmi, K., Seir, R. A., Abdellatief, I. & Abdeen, Z. Molecular Characterization
550 of Methicillin Resistant Staphylococcus aureus in West Bank-Palestine. *Front Public Health*
551 **7**, 130 (2019).
552 50. Monecke, S. *et al.* Molecular Typing of ST239-MRSA-III From Diverse Geographic
553 Locations and the Evolution of the SCCmec III Element During Its Intercontinental Spread.
554 *Front Microbiol* **9**, 1436 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

555 51. Alseqely, M. *et al.* Association between fluoroquinolone resistance and MRSA genotype in
556 Alexandria, Egypt. *Scientific Reports* (2021).
557 52. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to
558 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
559 53. Davis, J. J. *et al.* The PATRIC Bioinformatics Resource Center: expanding data and analysis
560 capabilities. *Nucleic Acids Res* **48**, D606–D612 (2020).
561 54. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**,
562 6614–6624 (2016).
563 55. Larsen, M. V. *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin*
564 *Microbiol* **50**, 1355–1361 (2012).
565 56. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics:
566 BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **3**,
567 124 (2018).
568 57. Bartels, M. D. *et al.* Comparing whole-genome sequencing with Sanger sequencing for spa
569 typing of methicillin-resistant Staphylococcus aureus. *J Clin Microbiol* **52**, 4305–4308
570 (2014).
571 58. Kondo, Y. *et al.* Combination of multiplex PCRs for staphylococcal cassette chromosome
572 mec type assignment: rapid identification system for mec, ccr, and major differences in
573 junkyard regions. *Antimicrob Agents Chemother* **51**, 264–274 (2007).
574 59. International Working Group on the Classification of Staphylococcal Cassette Chromosome
575 Elements (IWG-SCC). Classification of staphylococcal cassette chromosome mec
576 (SCCmec): guidelines for reporting novel SCCmec elements. *Antimicrob Agents Chemother*
577 **53**, 4961–4967 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.01.442118; this version posted May 1, 2021. 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.

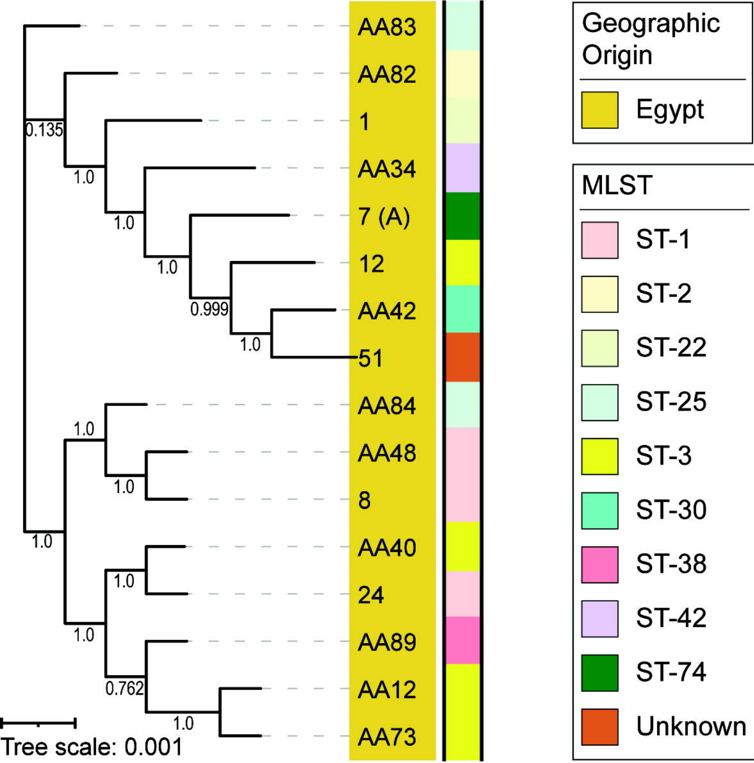
578 60. Wattam, A. R. *et al.* Improvements to PATRIC, the all-bacterial Bioinformatics Database
579 and Analysis Resource Center. *Nucleic Acids Res.* **45**, D535–D542 (2017).
580 61. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: a comparative pathogenomic
581 platform with an interactive web interface. *Nucleic Acids Res.* **47**, D687–D692 (2019).
582 62. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data.
583 *PeerJ* **3**, e1319 (2015).
584 63. Delmont, T. O. & Eren, A. M. Linking pangenomes and metagenomes: the Prochlorococcus
585 metapangenome. *PeerJ* **6**, e4320 (2018).
586 64. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome
587 coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.*
588 **43**, D261-269 (2015).
589 65. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
590 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
591 66. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood
592 Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
593 67. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new
594 developments. *Nucleic Acids Research* **47**, W256–W259 (2019).
595

a

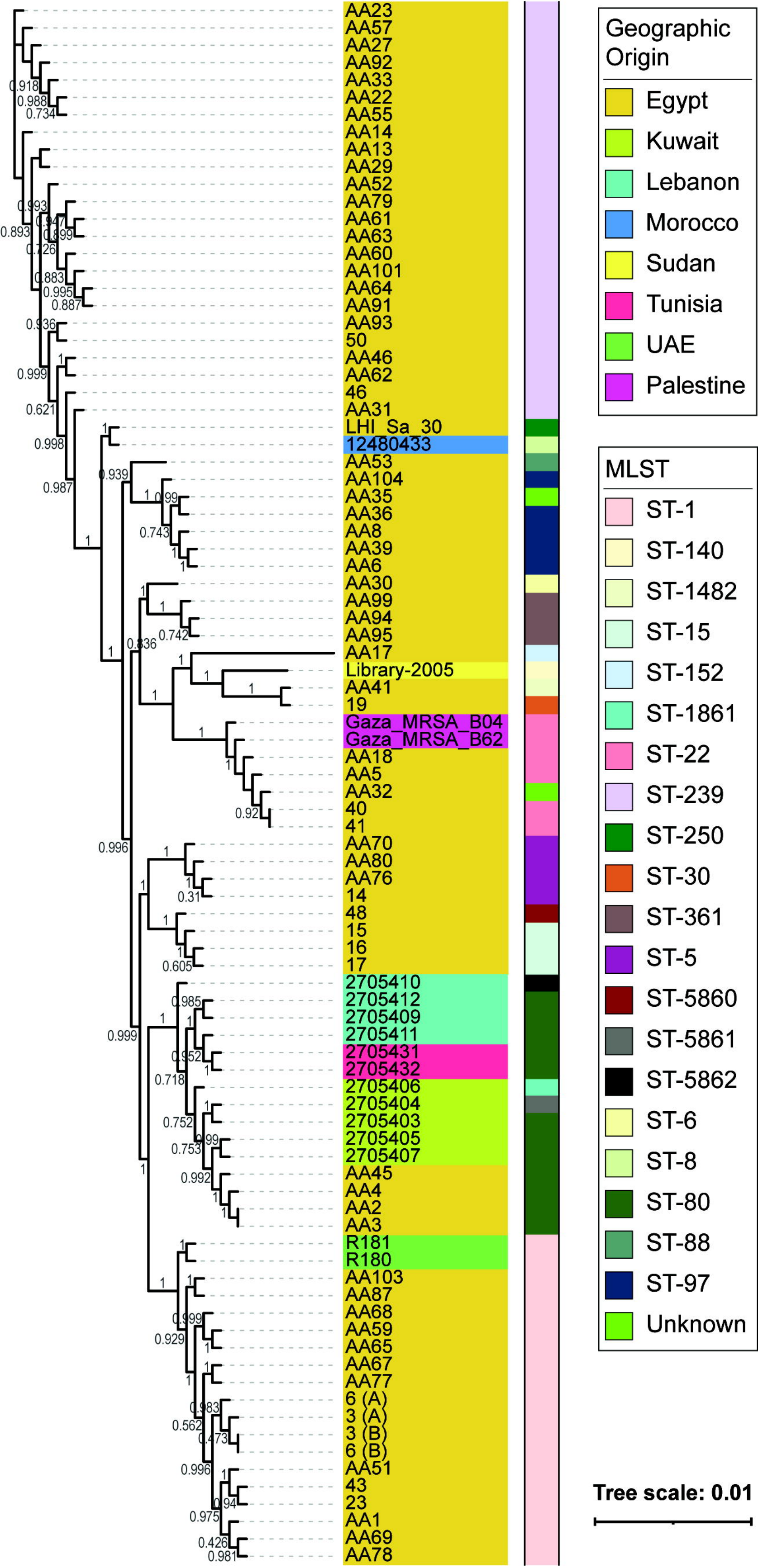
2822
Gene
Clusters
0
98
Singleton
Genes
0

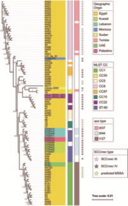
b

X: Mobilome: prophages, transposons
V: Defense mechanisms
B: Chromatin structure & dynamics
U: Intracellular trafficking, secretion, and vesicular transport
C: Energy production & conversion
T: Signal transduction mechanisms
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport & metabolism
S: Function unknown
F: Nucleotide transport & metabolism
G: Carbohydrate transport & metabolism
R: General function prediction only
H: Coenzyme transport & metabolism
Q: 2ndary metabolites biosynthesis, transport & catabolism
I: Lipid transport & metabolism
P: Inorganic ion transport & metabolism
O: Posttranslational modification, protein turnover, chaperones
J: Translation, ribosomal structure & biogenesis
M: Cell wall, membrane envelope biogenesis
K: Transcription
N: Cell motility
L: Replication, recombination & repair



AA83
AA82
1
AA34
7 (A)
12
AA42
51
AA84
AA48
8
AA40
24
AA89
AA12
AA73
0.135
1.0
1.0
1.0
0.999
1.0
1.0
1.0
1.0
1.0
1.0
0.762
1.0
Tree scale: 0.001
Geographic Origin
Egypt
MLST
ST-1
ST-2
ST-22
ST-25
ST-3
ST-30
ST-38
ST-42
ST-74
Unknown


AA23
AA57
AA27
AA92
AA33
AA22
AA55
AA14
AA13
AA29
AA52
AA79
AA61
AA63
AA60
AA101
AA64
AA91
AA93
50
AA46
AA62
46
AA31
LHI_Sa_30
12480433
AA53
AA104
AA35
AA36
AA8
AA39
AA6
AA30
AA99
AA94
AA95
AA17
Library-2005
AA41
19
Gaza_MRSA_B04
Gaza_MRSA_B62
AA18
AA5
AA32
40
41
AA70
AA80
AA76
14
48
15
16
17
2705410
2705412
2705409
2705411
2705431
2705432
2705406
2705404
2705403
2705405
2705407
AA45
AA4
AA2
AA3
R181
R180
AA103
AA87
AA68
AA59
AA65
AA67
AA77
6 (A)
3 (A)
3 (B)
6 (B)
AA51
43
23
AA1
AA69
AA78
Geographic Origin
Egypt
Kuwait
Lebanon
Morocco
Sudan
Tunisia
UAE
Palestine
MLST
ST-1
ST-140
ST-1482
ST-15
ST-152
ST-1861
ST-22
ST-239
ST-250
ST-30
ST-361
ST-5
ST-5860
ST-5861
ST-5862
ST-6
ST-8
ST-80
ST-88
ST-97
Unknown
Tree scale: 0.01


Geographic Origin
Egypt
Kuwait
Lebanon
Morocco
Sudan
Tunisia
UAE
Palestine
MLST CC
CC1
CC30
CC5
CC8
CC97
CC15
CC22
ST-80
spa type
t037
t044
t127
SCCmec type
SCCmec III
SCCmec IV
predicted MSSA
Tree scale: 0.01