1   **Genomic investigation of MRSA bacteremia relapse reveals diverse genomic**

2   **profiles but convergence in bacteremia-associated genes**

3   Brooke M. Talbot[1], Natasia F. Jacko[2,3], Katrina S. Hofstetter[1], Tara Alahakoon[2], Kevin

4   Bouiller[2,4], Timothy D. Read[1], Michael Z. David[2]#

5   **1.** Division of Infectious Diseases, Emory University School of Medicine, Atlanta, Georgia,

6   USA.

7   **2.** Division of Infectious Diseases, Department of Medicine, University of Pennsylvania,

8   Philadelphia, PA, USA.

9   **3.** Department of Biostatistics, Epidemiology and Informatics Perelman School of Medicine,

10   University of Pennsylvania, Philadelphia, PA, USA.

11   **4.** Université Marie et Louis Pasteur, CHU Besançon, CNRS, Chrono-environnement (UMR

12   6249), F-25000 Besançon, France.

13

14   #Correspondance should be directed to

15   Michael Z. David, M.D., Ph.D.

16   Division of Infectious Diseases, Department of Medicine, University of Pennsylvania

17   423 Guardian Drive, Room 707

18   Philadelphia, PA, USA 19104

19   E-mail: michdav@pennmedicine.upenn.edu

20   Alternate Contact: Timothy D. Read, 1750 Haygood Dr., Atlanta, GA, USA 30322

21   508-649-9023, tread@emory.edu

22   **Running head:** MRSA bacteremia relapse genomic profiles

23   **Word Count**: 3499/3500

Abstract

Background: Recurrence of methicillin-resistant *Staphylococcus aureus* (MRSA) bacteremia is a high risk complication for patients. Distinguishing persistent lineages from new infections is not standardized across clinical studies.

Methods: We investigated factors contributing to recurrence of MRSA bacteremia among subjects in Philadelphia, Pennsylvania. Subject demographics and clinical history were collected and paired with whole-genome sequences of infection isolates. Recurrent bacteremia episodes were recorded and defined as relapse infections (same lineage) or new infections by genomic criteria, where a relapse contains isolates <=25 single nucleotide polymorphisms (SNP) different, and by clinical criteria. All isolates were assessed for pairwise SNP distances, common mutations, and signatures of within-host adaptation using the McDonald-Kreitman test. Clusters of transmission between relapse-associated isolates and other subject lineages were identified.

Results: Among 411 sequential subjects with MRSA bacteremia, 32 experienced recurrent bacteremia episodes, with 24 subjects having exclusively relapse infections, six with infections exclusively from a new strain, and two patients with both relapse and new infections. No concordance between a genomic and a clinical definition of relapse was evident (Cohen's Kappa = 0.18, CI: -0.41). Recurrence-associated lineages exhibited signatures of positive selection(G-test:<0.01) . Genes with SNPs occurring in multiple relapse lineages have roles in antibiotic resistance and virulence, including 5 lineages with mutations in *mprF* and 3 lineages with mutations in *rpoB,* which corresponded with evolved phenotypic changes in daptomycin and rifampin resistance.

Conclusions: Recurrent infections have a diverse strain background. Relapses can be readily distinguished from newly acquired infections using genomic sequencing but not clinical criteria.

3

47   **Word Count:** 249/250

48   **Key words:** hospital epidemiology, genomic epidemiology, infection prevention, antibiotic

49   resistance

Introduction

Staphylococcus aureus bacteremia is a complex clinical syndrome which often leads to adverse patient outcomes, including endocarditis and other metastatic infections [1]. Recurrence of *S. aureus* bacteremia, when patients are diagnosed with bacteremia again after assumed resolution of a previous infection, is an ongoing clinical challenge. Global data demonstrate that 5 - 15 percent methicillin-resistant *S. aureus* (MRSA) bacteremia (MRSAB) episodes result in recurrence, and the risks associated with recurrence are similarly heterogenous to bloodstream infections overall [2–9]. Known risk factors for recurrence of *S. aureus* bacteremia in adults include younger patient age, presence of a foreign body, hemodialysis dependence, valvular heart disease, liver cirrhosis, and endocarditis [6,9].

Distinguishing new infections after clearance from cryptically persisting bacteria within host tissues is a challenge for clinicians, especially if the focus of an infection is unknown. We split recurrent MRSAB into two types: "relapses", caused by the same genetic lineage, and "new infections" in which a different *S. aureus* lineage is responsible. Recurrence has been heterogeneously defined using strain genotypes (e.g., *spa* typing or multilocus sequence typing [MLST]), time intervals, or clinical history of previous infections [4,6–9]. Without including genomic similarity criteria, it has not been possible to distinguish relapse from new infections accurately. Consequently, misidentification of the two types of recurrence includes, for relapses, not identifying persistent infections, and for new infections, that a patient may have a new risk factor for *S. aureus* infection. Defining the nature of a recurrence may help therapeutic approaches for preventing subsequent infections.

Colonization by *S. aureus* has been associated with higher risk of developing bacteremia. The transition from colonization to invasive infection, and the subsequent treatment regimen, can

73   result in bacterial adaptive evolution [10–12]. Mutations have been found to have occurred

74   convergently in multiple different MRSAB isolates. These genetic changes alter virulence

75   regulation [13–25] and antibiotic targets [13–15,26,27] and result in the development of small

76   colony variants [28,29] among other phenotypes. Emerging mutations can confer cross resistance

77   to first-line antibiotic treatments or multidrug resistance, further complicating prevention of

78   persistent infections or cure in patients with deep-seated foci [10]. Several loci appear to increase

79   the risk for metastatic and persistent infections, and possible recurrence. An array of putative loci

80   have been identified to adapt during *S. aureus* bacteremia, but it is not known which, if any, are

81   more associated with relapse versus new infection. There is also little information on how *S.*

82   *aureus* strain background influences forward evolution during bacteremia or the propensity for

83   relapse. Knowledge of genetic change patterns in *S. aureus* during bacteremia episodes may help

84   differentiate new infections from relapses.

85   To define genetic differences and risks for MRSAB recurrence, we examined a cohort of

86   patients in Philadelphia, Pennsylvania experiencing MRSAB during a five-year period who

87   received care at a single hospital system. We performed whole-genome sequencing (WGS) on

88   single isolates from single MRSAB episodes and recurrent episodes. We examined the

89   associations between host clinical factors and bacterial genetics to characterize recurrent

90   MRSAB, and defined the transmission and adaptive history of infections that resulted in a

91   relapse.

Methods

92

Subject cohort and isolate collection

93

94   This study was considered exempt by the University of Pennsylvania Institutional

95   Review Board. Subject MRSAB isolates were included from a cohort of adult patients admitted

96   to at least one of two hospitals in Philadelphia, Pennsylvania and diagnosed with MRSAB

97   between July 2018 and February 2022. A single colony representative isolate of the first positive

98   culture was stored for each MRSAB episode. Clinical and demographic information was

99   collected through medical record reviews for all infections, including age at time of isolate

100   collection, race, ethnicity, sex at birth, death within 30 days of infection diagnosis,

101   comorbidities, complications from bacteremia, antibiotic treatment, suspected source site of

102   infection, and healthcare-associated epidemiological category. Antibiotic resistance phenotypes

103   were collected from clinical microbiology records associated with the unique MRSA isolates.

104   Antibiotic resistance was assessed using the Vitek 2 automated system, and assigned

105   susceptibility/resistance in accordance with Clinical and Laboratory Standards Institute protocols

106   [30].

107   ***Clinical distinction of relapse and new infections***

108   Subsequent MRSAB events per subject (a "recurrent episode") were classified as a

109   "relapse" or "new infection" according to the following criteria as outlined in Figure S1:

110   MRSAB episodes include all isolates collected in a 30 day period from the first episode isolate.

111   At each discrete medical encounter during which MRSAB was detected, subjects with one or

112   more MRSA isolate from blood were assessed for previous MRSAB and subdivided into discrete

113   bacteremia episodes if greater than 30 days apart. Subjects with multiple discrete bacteremia

114   episodes were considered to have one or more "recurrent episodes." Recurrent episodes were

115   then categorized into "new infections" and "relapse" infections using clinical criteria or genomic

116   criteria. For clinical criteria, new infections had to fulfill all of the following: The first positive

117   culture from the episode was 30 days or more following the last positive MRSA blood culture,

118   all symptoms at the source site and metastatic sites of the previous infection were resolved, no

119   new antibiotics were prescribed after completion of definitive therapy for the index MRSAB

120   episode aside from antibiotics aimed at suppression of an infected focus, if the subject was

121   administered suppressive antibiotics the infection site (source) was different from the previous

122   infection, if a central venous catheter was exchanged at the time of the previous MRSAB then it

123   was not exchanged over a wire but instead was resited [31], or the infection source was clinically

124   ruled to be different from the previous infection. Otherwise, if any one or more of these criteria

125   were not true the bacteremia episode was considered a relapse.

126   ***Sequencing quality and phylogenetics***

127   Genomic DNA was extracted from *S. aureus* isolates and sequenced at the

128   Penn/Children's Hospital of Philadelphia Microbiome Center, using paired-end short read

129   whole-genome sequencing as previously described [32]. Reads were processed using Bactopia (v

130   3.0.0) [33]. Briefly, adaptor sequences were removed, and reads were de novo assembled using

131   Shovill (v.1.1.0). Sequences were used for further investigation if reads had at least an average

132   per-read quality score of Q12 and a mean read length of 49bp, and the genomic assembly had at

133   least 20x coverage and no more than 500 contigs. Multilocus sequence type (ST) and clonal

134   complex (CC) were assigned by calling Ariba (v 2.14.6) within Bactopia, which utilized the *S.*

135   *aureus* specific ST scheme from PubMLST [34] . For novel STs or STs without a defined CC,

136   the CC was manually defined using a maximum likelihood (ML) tree and relative position to the

137   next isolate with a defined CC. A core genome alignment was created among all MRSAB

138   isolates and MSSA Strain Newman (GCF_020985245.1) as an outgroup with the Bactopia

139   subworkflow "pangenome" using PIRATE [35]. Areas of likely recombination were identified

140   and masked using ClonalFrame ML (v.1.12) [36]. A ML tree of all isolates was created from the

141   masked alignment with IQTree (2.1.2) [37] and ModelFinder [38], which determined the best fit

142   model to be a generalized time reversible model with Empirical base frequencies plus the

143   FreeRate model.

Genomic criteria for relapsed and new infections and cluster identification

145   Snp-dists (0.8.2) was used to calculate single nucleotide polymorphism (SNP) pairwise

146   distances between aligned isolates. Subjects with recurrent MRSAB episodes were categorized

147   as having a relapsed or new infection based on pairwise SNP differences from the isolate of the

148   most recent previous episode relative to the next subsequent episode. A conservative SNP

149   threshold of $<=25$ SNPs between episodes defined a relapse based on previous observations of

150   SNP distances of MRSA lineages persisting within a host [39]. The phylogenetic branch patterns

151   were also investigated between recurrent episode isolates: relapse isolates must also share their

152   most-recent common ancestor with another isolate from within the same human subject and

153   share the same ST. Cohen's kappa [40,41] was calculated between the clinical and genomic

154   definitions of relapse to assess method agreement and clinical criteria sensitivity and specificity

155   relative to the genomic criteria.

Association of clusters with demographic and clinical data

157   Demographic and clinical characteristics were compared between subjects with recurrent

158   MRSAB episodes and single MRSAB episodes using Kruskal-Wallis test for continuous

159   variables and Peason's Chi-square test or Fisher's exact test for categorical variables.

160   Demographic and clinical history were first compared between MRSAB subjects who died

161   within their index MRSAB episode window (30 days) and those who survived their index

162 infection (Table S1). Decedents from index MRSAB infection were excluded from the analysis

163 assessing risk factors associated with MRSAB infection survival.

164 Genomic similarity between subjects' isolates was also compared to identify potential

165 transmission clusters. For all relapse-associated isolates <= 25 SNPs from different subjects'

166 isolates, subtrees of the larger ML tree were created and examined for branch structure and days

167 since the earliest collection date in the cluster to identify the role of relapse isolates in putative

168 onward transmission.

169 ***Variant calling in relapse lineages***

170 Bactopia's Snippy subworkflow, which utilizes Snippy v4.6.0 [42], was used for variant

171 calling of SNPs of all isolates against a previously generated ancestrally reconstructed *S. aureus*

172 genome [43]. Variant calling was also conducted for each set of identified relapse infections by

173 using the first known isolate (index) in each lineage as the reference. Index isolates were

174 annotated using Bakta (v1.9.3)[44] to generate the reference sequence for variant analysis. The –

175 snippy-core option was run on all isolates to identify variants in core coding regions across all

176 genomes to produce an alignment and predicted changes to amino acid structure. To identify

177 commonly occurring non-synonymous changes within each recurrence lineage, the –snippy-core

178 option was also used to create a SNP alignment for each set of isolates from the same subject to

179 identify common variants and changes to amino acid structure. SNPs in coding regions were

180 concatenated and summarized for all lineages.

181 ***Creating a database of bacteremia-associated mutations***

182 A literature review was conducted to create a database of previously identified genes

183 and/or mutations in *S. aureus* genomes associated with bacteremia. PubMed was searched in

184 December 2023 using the query (((staphylococcus aureus)) AND ((bacteremia) OR (bloodstream

185   infection))) AND (genetic mutation). Only peer-reviewed articles were scanned for evidence.

186   Genes or mutations in genes were considered if the study reported that they occurred in the *S.*

187   *aureus* genome, samples were derived from *S. aureus* bacteremia, were associated with

188   characteristics relevant to an outcome (e.g., virulence expression or host survival) or were

189   associated with host phenotypic changes related. A listing of citations identifying these genes is

190   in Table S2.

191   *Calculation of the index of neutrality*

192   McDonald-Kreitman tests [45] were conducted on all MRSAB isolates and the subset of

193   relapse-associated isolates to assess selection on the whole genome or on known bacteremia-

194   associated genes. A core genome alignment was created using snippy-core for all isolates and for

195   isolates associated with a relapse lineage. The ancestral reference served as the outgroup. Fixed

196   sites were determined by counting the nucleotide sites universally conserved from the reference

197   among compared isolates. Polymorphic sites were determined if there were nucleotide sites that

198   differed between individuals within the isolate groups. A G-test was used to evaluate the

199   significance of the neutrality index.

200   # Results

201

202   **Recurrent MRSAB are from common clinical and phylogenetic backgrounds among**

203   **circulating strains.** We sequenced 456 *S. aureus* isolates from episodes of MRSAB and

204   included 411 subjects, of whom 32 had at least 1 recurrent MRSAB episode in addition to the

205   index infection (77 isolates). Recurrence-associated isolates occurred across major CCs of the *S.*

206   *aureus* species, including CC5 (N = 29), CC8 (N = 41), CC30 (N = 3), CC78 (N = 3) and CC72

207 (N = 1) (Fig. 1). CC distribution was similar between recurrence and non-recurrence isolates

208 (p=0.91), and there was no difference in the number of recurrences over time.

209 Subjects surviving the index MRSAB episode (N=311) were assessed for risk factors for

210 recurrent MRSAB (Table 1). When subjects with recurrent episodes were compared to those

211 with no record of recurrence, we found no significant differences in clinical or demographic

212 characteristics. Subjects epidemiologically categorized as having healthcare-acquired community

213 onset (HACO) infections were more frequent among recurrent subjects (84%) compared to non-

214 recurrence patients (66%), although these associations were not statistically significant.

215

216 **New infections are genetically distinct from relapse infections among recurrent episodes.**

217 Using genomic criteria we classified 24 subjects as having only relapse infections and six with

218 infections only from a new strain; two subjects experienced both relapse and new infections. We

219 identified 45 subsequent recurrent episodes among the 32 subjects. Relapse episodes occurred

220 sooner after the prior infection episode (N = 37, median 97 days, interquartile range, IQR 69-200

221 days) compared to new strain episodes (N = 8, median 311 days, IQR 79-568 days), though this

222 was not statistically significant (P=0.17) (Fig. 2A). Most relapses fell well below the 25 SNP

223 threshold, with only 3 episodes defined instead by their shared common ancestor with

224 intrasubject isolates. All isolates classified as new infections were distantly related to the

225 subject's previous infection and frequently belonging to a separate clade of *S. aureus* entirely

226 (Fig. 2B).

227 When clinical and genomic definitions were compared, the overall concordance was poor

228 (Cohen's Kappa = 0.18, CI: -0.41), with the genomic definition predicting that 82% of

229 subsequent infections were related to the previous infection, and the clinical definition predicting

230   that 50% were related (Fig. 3). Using genomics as a gold standard for relapse, the clinical

231   definition had a sensitivity of 55% and a specificity of 75%. When a device or foreign body was

232   implicated in any infection, however, there was often high concordance identifying relapse

233   infections. The genomic definitions of relapse and new infections were used for the remainder of

234   analyses.

235   **Relapse infections demonstrate distinct adaptation to the host**. Across all isolates, there was

236   a trend of neutral evolution across the genome and in the subset of genes associated with

237   bacteremia. Comparatively, relapse-associated lineages showed a significant signature of

238   positive selection in the whole genome (Table 2). Although relatively few bacteremia-associated

239   genes were synonymously mutated, there were notably no synonymous mutations at

240   polymorphic sites. Together, this suggests that relapse-associated lineages likely undergo

241   ongoing selection after the index infection and dissemination.

242
243   **Antibiotic resistance genotypes and phenotypes in relapses correspond with patient**

244   **exposures**. Eleven genes had unique SNPs with mutations in two or more relapse lineages (Fig.

245   4A). Mutations in these genes occurred regardless of CC, indicating that clade background alone

246   did not contribute to mutations in these genes. Genes in which multiple relapse lineages showed

247   non-synonymous mutations were implicated in known virulence traits and antibiotic resistance.

248   The genes most commonly mutated were *mprF*, in five separate subject lineages, and *rpoB,* in

249   four subject lineages. Changes in both *mprF* and *rpoB* are associated with drug resistance to

250   antibiotics commonly used for bacteremia; we therefore investigated amino acid impact, isolate

251   MIC, and treatment course. Multiple amino acid changes were detected between patients for

252   proteins encoded by both genes. The same amino acid change, Ala477Asp, occurred for *rpoB*

253   products in two subjects with rifampin resistance; for one subject the resistance phenotype was

254   present in their first infection before relapse, a change at position 477 and position 471

255   (Asp471Tyr), which corresponded with a phenotypic change from susceptible to resistant (Fig.

256   4B). Three subjects with *mprF* mutations emerging in relapses demonstrated acquired

257   daptomycin resistance with changes at Ser337Thr, Ser337Leu, and Leu291Ile (Fig. 4C). All

258   recurrence-associated subjects with *mprF* mutations had exposure to daptomycin prior to

259   mutation emergence regardless of daptomycin resistance phenotype. No other relapse lineages

260   gained or lost resistance to rifampin or daptomycin without emergence of *mprF* or *rpoB*

261   mutations.

262   **Relapse isolates cluster with other subject isolates, but do not contribute to onward**

263   **transmission**. Nine clusters across subjects including at least one subject who had a relapse were

264   identified. These clusters occurred in distinct lineages across CC5 and CC8 clades (Fig. 5). If

265   relapse infections were contributing to onward transmission, or if patients were reinfected with a

266   closely related circulating strain, we might expect that infections in unique hosts would cluster

267   within relapse lineage clades. Across all nine clusters, isolates from different subjects clustered

268   significantly outside of the relapse lineage isolates. This suggested that transmission occurred

269   before the onset of relapsing infection and the unique lineages within a host were highly specific

270   to the individual.

271

Discussion

273   Our study showed that recurrent MRSAB is well differentiated into new and relapse

274   infections by WGS and patterns of pathogen evolution. Consensus of a time and genomic

275   definition of relapse is still not standard. It is imperative to document and compare genotypic and

276   phenotypic patters among recurrent isolates that accompany different molecular definitions. For

277   our investigation period shows that most isolates from the same person were separated by fewer

278   than 25 core genome SNPs. We defined 3 additional isolates from episodes as relapses upon

279   inspection of their phylogenetic relatedness to intrahost isolates. Previous studies have identified

280   a similar pattern in distinguishing relapses from new infections even using less precise

281   technologies including pulsed-field gel electrophoresis finding that approximately an equal

282   number were classified as relapse and new infections [9]. Our SNP-based definition revealed a

283   much larger percentage of relapse infections than previous studies would suggest. We observed

284   unique phenotypic traits (i.e., daptomycin resistance) within relapse lineages, which would

285   suggest emergence and likely persistence of a strain within the host. Though external reservoirs

286   cannot be entirely ruled out, the nesting of the infections and the accumulation of genetic

287   changes suggest a high likelihood of specific host-association. Furthermore, while phylogenetic

288   histories of relapse-associated MRSAB isolates do occur in transmission clusters, they do not

289   appear to be directly contributing to ongoing spread leading to additional MRSAB cases in

290   healthcare settings.

291   Previous studies have identified risk factors for *S. aureus* bacteremia and MRSAB

292   recurrence related to pathogen and host factors, and the type of care received. We identified

293   similar risk factors in our dataset to those recorded previously, including a high prevalence of

294   SCC*mec* type II [8] and a high proportion of unremoved foreign objects [6,7]. We did not

295   identify a significant association between antibiotic treatment duration and recurrence. Other

296   reports show a negative association between antibiotic duration and relapse [4,45]. These

297   discrepancies may be explained by the type and duration of antibiotic treatment. Nevertheless,

298   the number of recurrent infections we detected is from a single hospital group and may not be

299   generalizable to other populations. Larger population sizes will be needed to identify variables

300   with small effect sizes.

301   We found evidence that antibiotic therapy may have selected for strains with mutations in

302   genes known to confer daptomycin and rifampin resistance. Several distinct amino acid changes

303   were detected in isolates with both rifampin-resistant and daptomycin-resistant phenotypes at the

304   index infection and acquired in some, as demonstrated in relapse isolate genomes. Changes to the

305   amino acid sequence corresponded with expected phenotypic changes observed in other studies

306   such as the changed asparagine at position 481 of the *rpoB* transcription product resulting in a

307   rifampin-intermediate resistance profile [47], and the point mutations observed in *mprF* that

308   resulted in changes at the same amino acid site (Ser337Thr and Ser337Leu) corresponding to

309   daptomycin resistance. Multiple amino acid changes corresponding with phenotypic changes

310   here demonstrate wide diversity of mutational profiles of *rpoB* and *mprF* even within a single

311   study population and convergence of traits within infected patients rather than single, resistant

312   strains spreading between individuals.

313   Management of MRSAB involves a comprehensive assessment of patient history,

314   physical examination, and source identification [1]. Delays removing or draining an infection

315   focus can increase the risk of persistent bacteremia [1] or metastatic spread [48]. When central

316   venous catheters or other foreign bodies are suspected to be the source of SAB, removal of the

317   foreign body is considered, though it is not always possible when it increases risk of patient

318   harm. We found genomically similar MRSAB episodes common among patients with foreign

319   body infections, suggesting the presence of persistent infection foci reseeding the blood. We had

320   previously reported MRSA infections associated with a foreign body have a nearly 5-fold greater

321   odds among patients who had a previous MRSA infection within a year compared to those with

322   no previously reported infection [49]. Clinicians should maintain a high index of suspicion of

323   implanted foreign bodies as a source of recurrent MRSAB and advise patients accordingly

324   during MRSAB recovery.

325   WGS provided additional support for case-identification of new and relapsing infections

326   in the context of the clinical history. The complexity of host factors, within-host selection

327   pressures, strain background and type of antibiotic treatment all play interacting roles in relapse

328   emergence. It is not be possible to determine if recurrent MRSAB is caused by the same strain

329   without performing WGS. Examining isolates in a population context further differentiated

330   relapses from new infections. Relapsing isolates undergo positive adaptation to the host, and

331   convergently mutating genes are consistent with long-term usage or high exposure to antibiotics,

332   but other traits necessary for survival in the body may still play an important role in persistence.

333   Additional work is necessary to understand survival duration of *S. aureus* strains across

334   individual hosts' body sites. Combining the frequency of genetic mutations, genetic relatedness,

335   and known clinical risk factors for recurrence will lay the groundwork for better prediction of

336   new MRSAB and other *S. aureus* bacteremia relapses and may contribute to therapy strategies to

337   prevent relapsing infections.

338

339

340

341

342

Acknowledgements

Author contributions

BMT: Conceptualized project, designed analysis, collected data, analyzed data, drafted initial manuscript, reviewed and approved manuscript.

NFJ: collected data, reviewed and approved manuscript.

KH: collected data, reviewed analysis, reviewed and approved manuscript.

TA: Collected data, reviewed and approved manuscript.

KB: Collected data, reviewed and approved manuscript.

MZD: Conceptualized project, designed analysis, collected data, drafted the initial manuscript, reviewed and approved manuscript.

TDR: Conceptualized project, designed analysis, drafted the initial manuscript, reviewed and approved manuscript.

Data availability

Raw sequence reads for this study are publicly available in the Sequence Read Archive under the project ID PRJNA751847. Metadata and analysis code used to generate the figures and tables are available at https://doi.org/10.5281/zenodo.15066414

Funding

370

Conflicts of Interest

372   The authors report no conflicts of interest.

373

References

375   1.   Minter DJ, Appa A, Chambers HF, Doernberg SB. Contemporary Management of

376   Staphylococcus aureus Bacteremia-Controversies in Clinical Practice. Clin Infect Dis Off

377   Publ Infect Dis Soc Am. United States; **2023**; 77(11):e57–e68.

378   2.   Holmes NE, Turnidge JD, Munckhof WJ, et al. Genetic and molecular predictors of high

379   vancomycin MIC in Staphylococcus aureus bacteremia isolates. J Clin Microbiol. United

380   States; **2014**; 52(9):3384–3393.

381   3.   Small PM, Chambers HF. Vancomycin for Staphylococcus aureus endocarditis in

382   intravenous drug users. Antimicrob Agents Chemother. United States; **1990**; 34(6):1227–

383   1231.

384 4. Raad I, Narro J, Khan A, Tarrand J, Vartivarian S, Bodey GP. Serious complications of

385   vascular catheter-related Staphylococcus aureus bacteremia in cancer patients. Eur J Clin

386   Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. Germany; **1992**; 11(8):675–682.

387 5. Ehni WF, Reller LB. Short-course therapy for catheter-associated Staphylococcus aureus

388   bacteremia. Arch Intern Med. United States; **1989**; 149(3):533–536.

389 6. Choi S-H, Dagher M, Ruffin F, et al. Risk Factors for Recurrent Staphylococcus aureus

390   Bacteremia. Clin Infect Dis. **2021**; 72(11):1891–1899.

391 7. Fowler VGJ, Kong LK, Corey GR, et al. Recurrent Staphylococcus aureus bacteremia:

392   pulsed-field gel electrophoresis findings in 29 patients. J Infect Dis. United States; **1999**;

393   179(5):1157–1161.

394 8. Welsh KJ, Skrobarcek KA, Abbott AN, et al. Predictors of relapse of methicillin-resistant

395   Staphylococcus aureus bacteremia after treatment with vancomycin. J Clin Microbiol.

396   United States; **2011**; 49(10):3669–3672.

397 9. Chang F-Y, MacDonald BB, Peacock JEJ, et al. A prospective multicenter study of

398   Staphylococcus aureus bacteremia: incidence of endocarditis, risk factors for mortality, and

399   clinical impact of methicillin resistance. Medicine (Baltimore). United States; **2003**;

400   82(5):322–332.

401 10. Giulieri SG, Guérillot R, Duchene S, et al. Niche-specific genome degradation and

402   convergent evolution shaping Staphylococcus aureus adaptation during severe infections.

403   eLife. England; **2022**; 11.

404   11. Culyba MJ, Van Tyne D. Bacterial evolution during human infection: Adapt and live or
405   adapt and die. PLoS Pathog. United States; **2021**; 17(9):e1009872.

406   12. Golubchik T, Batty EM, Miller RR, et al. Within-Host Evolution of Staphylococcus aureus
407   during Asymptomatic Carriage. PLOS ONE. Public Library of Science; **2013**; 8(5):e61319.

408   13. Gao W, Chua K, Davies JK, et al. Two novel point mutations in clinical Staphylococcus
409   aureus reduce linezolid susceptibility and switch on the stringent response to promote
410   persistent infection. PLoS Pathog. United States; **2010**; 6(6):e1000944.

411   14. Gao W, Cameron DR, Davies JK, et al. The RpoB H□□□Y rifampicin resistance mutation
412   and an active stringent response reduce virulence and increase resistance to innate immune
413   responses in Staphylococcus aureus. J Infect Dis. United States; **2013**; 207(6):929–939.

414   15. Giulieri SG, Baines SL, Guerillot R, et al. Genomic exploration of sequential clinical
415   isolates reveals a distinctive molecular signature of persistent Staphylococcus aureus
416   bacteraemia. Genome Med. England; **2018**; 10(1):65.

417   16. Hachani A, Giulieri SG, Guérillot R, et al. A high-throughput cytotoxicity screening
418   platform reveals agr-independent mutations in bacteraemia-associated Staphylococcus
419   aureus that promote intracellular persistence. eLife. England; **2023**; 12.

420   17. Beenken KE, Mrak LN, Zielinska AK, et al. Impact of the functional status of saeRS on in
421   vivo phenotypes of Staphylococcus aureus sarA mutants. Mol Microbiol. England; **2014**;
422   92(6):1299–1312.

423   18.   Liu Y, Gao W, Yang J, Guo H, Zhang J, Ji Y. Contribution of Coagulase and Its Regulator

424   SaeRS to Lethality of CA-MRSA 923 Bacteremia. Pathog Basel Switz. Switzerland; **2021**;

425   10(11).

426   19.   Benoit JB, Frank DN, Bessesen MT. Genomic evolution of Staphylococcus aureus isolates

427   colonizing the nares and progressing to bacteremia. PloS One. United States; **2018**;

428   13(5):e0195860.

429   20.   Bryson D, Hettle AG, Boraston AB, Hobbs JK. Clinical Mutations That Partially Activate

430   the Stringent Response Confer Multidrug Tolerance in Staphylococcus aureus. Antimicrob

431   Agents Chemother. United States; **2020**; 64(3).

432   21.   Nygaard TK, Pallister KB, Ruzevich P, Griffith S, Vuong C, Voyich JM. SaeR binds a

433   consensus sequence within virulence gene promoters to advance USA300 pathogenesis. J

434   Infect Dis. United States; **2010**; 201(2):241–254.

435   22.   Altman DR, Sullivan MJ, Chacko KI, et al. Genome Plasticity of agr-Defective

436   Staphylococcus aureus during Clinical Infection. Infect Immun. United States; **2018**;

437   86(10).

438   23.   Tsuji BT, Harigaya Y, Lesse AJ, Sakoulas G, Mylotte JM. Loss of vancomycin bactericidal

439   activity against accessory gene regulator (agr) dysfunctional Staphylococcus aureus under

440   conditions of high bacterial density. Diagn Microbiol Infect Dis. United States; **2009**;

441   64(2):220–224.

442 24. Cheung AL, Eberhardt KJ, Chung E, et al. Diminished virulence of a sar-/agr- mutant of

443   Staphylococcus aureus in the rabbit model of endocarditis. J Clin Invest. United States;

444   **1994**; 94(5):1815–1822.

445 25. Chong YP, Kim ES, Park S-J, et al. Accessory gene regulator (agr) dysfunction in

446   Staphylococcus aureus bloodstream isolates from South Korean patients. Antimicrob

447   Agents Chemother. United States; **2013**; 57(3):1509–1512.

448 26. Villar M, Marimón JM, García-Arenzana JM, Campa AG de la, Ferrándiz MJ, Pérez-

449   Trallero E. Epidemiological and molecular aspects of rifampicin-resistant Staphylococcus

450   aureus isolated from wounds, blood and respiratory samples. J Antimicrob Chemother.

451   England; **2011**; 66(5):997–1000.

452 27. Bæk KT, Thøgersen L, Mogenssen RG, et al. Stepwise decrease in daptomycin

453   susceptibility in clinical Staphylococcus aureus isolates associated with an initial mutation

454   in rpoB and a compensatory inactivation of the clpX gene. Antimicrob Agents Chemother.

455   United States; **2015**; 59(11):6983–6991.

456 28. Souza DC de, Cogo LL, Palmeiro JK, et al. Thymidine-auxotrophic Staphylococcus aureus

457   small-colony variant bacteremia in a patient with cystic fibrosis. Pediatr Pulmonol. United

458   States; **2020**; 55(6):1388–1393.

459 29. Douglas EJA, Duggan S, Brignoli T, Massey RC. The MpsB protein contributes to both the

460   toxicity and immune evasion capacity of Staphylococcus aureus. Microbiol Read Engl.

461   England; **2021**; 167(10).

462   30.   CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 29th ed. CLSI

463   supplement M100. Wayne, PA: Clinical and Laboratory Standards Institute; 2019.

464   31.   David MZ, Daum RS. Treatment of Staphylococcus aureus Infections. In: Bagnoli F,

465   Rappuoli R, Grandi G, editors. Staphylococcus Aureus Microbiol Pathol Immunol Ther

466   Prophyl [Internet]. Cham: Springer International Publishing; 2017. p. 325–383. Available

467   from: https://doi.org/10.1007/82_2017_42

468   32.   Talbot BM, Jacko NF, Petit III RA, et al. Unsuspected clonal spread of methicillin-resistant

469   Staphylococcus aureus causing bloodstream infections in hospitalized adults detected using

470   whole genome sequencing. Clin Infect Dis. Oxford University Press US; **2022**;

471   75(12):2104–2112.

472   33.   Petit Robert A., Read Timothy D., Segata Nicola. Bactopia: a Flexible Pipeline for

473   Complete Analysis of Bacterial Genomes. mSystems. American Society for Microbiology;

474   5(4):e00190-20.

475   34.   Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb

476   software, the PubMLST.org website and their applications. Wellcome Open Res. England;

477   **2018**; 3:124.

478   35.   Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. PIRATE: A fast and scalable

479   pangenomics toolbox for clustering diverged orthologues in bacteria. GigaScience. **2019**;

480   8(10):giz119.

481   36.   Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole

482   Bacterial Genomes. PLOS Comput Biol. Public Library of Science; **2015**; 11(2):e1004041.

483   37.   Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient

484   Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. **2020**; 37(5):1530–

485   1534.

486   38.   Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. ModelFinder: fast

487   model selection for accurate phylogenetic estimates. Nat Methods. **2017**; 14(6):587–589.

488   39.   Read TD, Petit RA, Yin Z, Montgomery T, McNulty MC, David MZ. USA300

489   Staphylococcus aureus persists on multiple body sites following an infection. BMC

490   Microbiol. **2018**; 18(1):206.

491   40.   Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. Sage

492   Publications Sage CA: Thousand Oaks, CA; **1960**; 20(1):37–46.

493   41.   Hazra A, Gogtay N. Biostatistics Series Module 7: The Statistics of Diagnostic Tests.

494   Indian J Dermatol. India; **2017**; 62(1):18–24.

495   42.   Seemann T. snippy: fast bacterial variant calling from NGS reads [Internet]. 2015.

496   Available from: https://github.com/tseemann/snippy

497   43.   Joseph SJ, Li B, Petit Iii RA, Qin ZS, Darrow L, Read TD. The single-species metagenome:

498   subtyping Staphylococcus aureus core genome sequences from shotgun metagenomic data.

499   PeerJ. United States; **2016**; 4:e2571.

500   44.   Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid

501   and standardized annotation of bacterial genomes via alignment-free sequence

502   identification. Microb Genomics. England; **2021**; 7(11).

503   45. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila.

504   Nature. England; **1991**; 351(6328):652–654.

505   46. Abbas M, Rossel A, Kraker MEA de, et al. Association between treatment duration and

506   mortality or relapse in adult patients with Staphylococcus aureus bacteraemia: a

507   retrospective cohort study. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect

508   Dis. England; **2020**; 26(5):626–631.

509   47. Watanabe Yukiko, Cui Longzhu, Katayama Yuki, Kozue Kishii, Hiramatsu Keiichi. Impact

510   of rpoB Mutations on Reduced Vancomycin Susceptibility in Staphylococcus aureus. J Clin

511   Microbiol. American Society for Microbiology; **2020**; 49(7):2680–2684.

512   48. Horino T, Hori S. Metastatic infection during Staphylococcus aureus bacteremia. J Infect

513   Chemother. **2020**; 26(2):162–169.

514   49. Bouiller K, Jacko NF, Shumaker MJ, Talbot BM, Read TD, David MZ. Factors associated

515   with foreign body infection in methicillin-resistant Staphylococcus aureus bacteremia.

516   Front Immunol. Switzerland; **2024**; 15:1335867.

517

518

519

520

521

522

523

524 # Figures
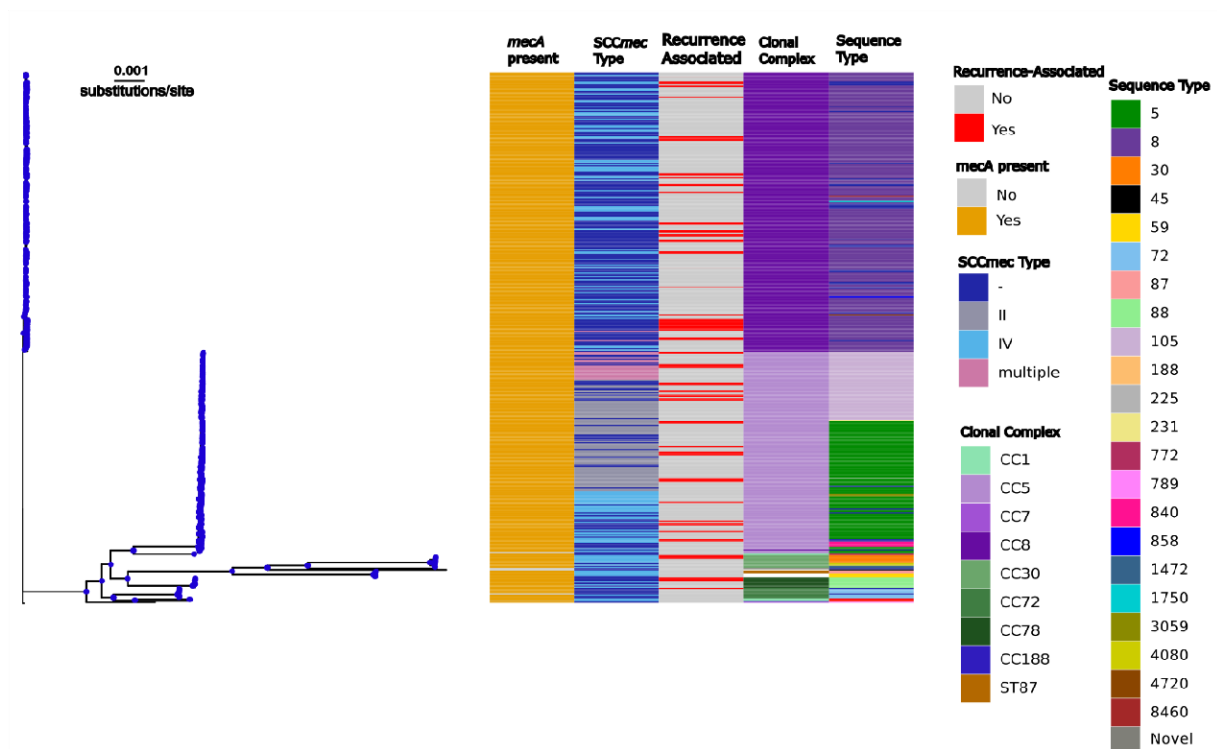
525
526 **Figure 1: Recurrent MRSAB isolates share a similar phylogenetic distribution to non-**

527 **recurrent MRSAB episodes.** A core pangenome tree of 411 MRSA bloodstream isolates was

528 constructed and rooted using MSSA strain Newman (GCF_020985245.1). Whether or not an

529 isolate was associated with a recurrent episode and molecular characteristics are shown in the

530 heat map including, from left to right, the presence or absence of *mecA*, *mecA* type, clonal

531 complex, and mutlilocus sequence type (ST). Blue dots indicated nodes where Shimodaira-

532 Hasegawa approximate likelihood ratio test and ultrafast bootstrap values were >70).

533
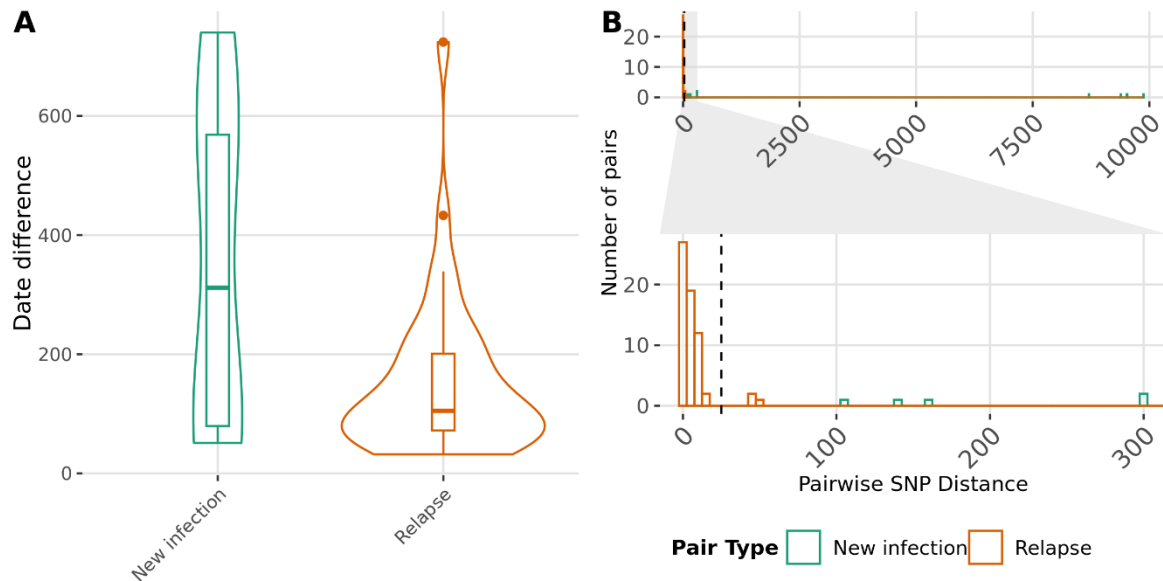
Figure 2: Relapsing infections and new infections within a patient are genomically

distinguishable. Subsequent isolate pairs among recurrences episodes (n=45) were separated

into relapse-associated (n=37) or new infections (n=8) based on pairwise single nucleotide

polymorphism (SNP) distance and distance to last common ancestor between isolates within the

subject (A) Difference in time between subsequent episodes for genomically new infections and

relapse associated infections were compared using the Kruskal-Wallice test (p=0.17). (B) Counts

of the pairwise SNP distances between all isolate pairs from the same host were quantified

(N=73). The inset display demonstrates counts where the SNP distance between isolates was

between 0 and 350 SNPs. The dotted line represents the 25 SNP threshold.
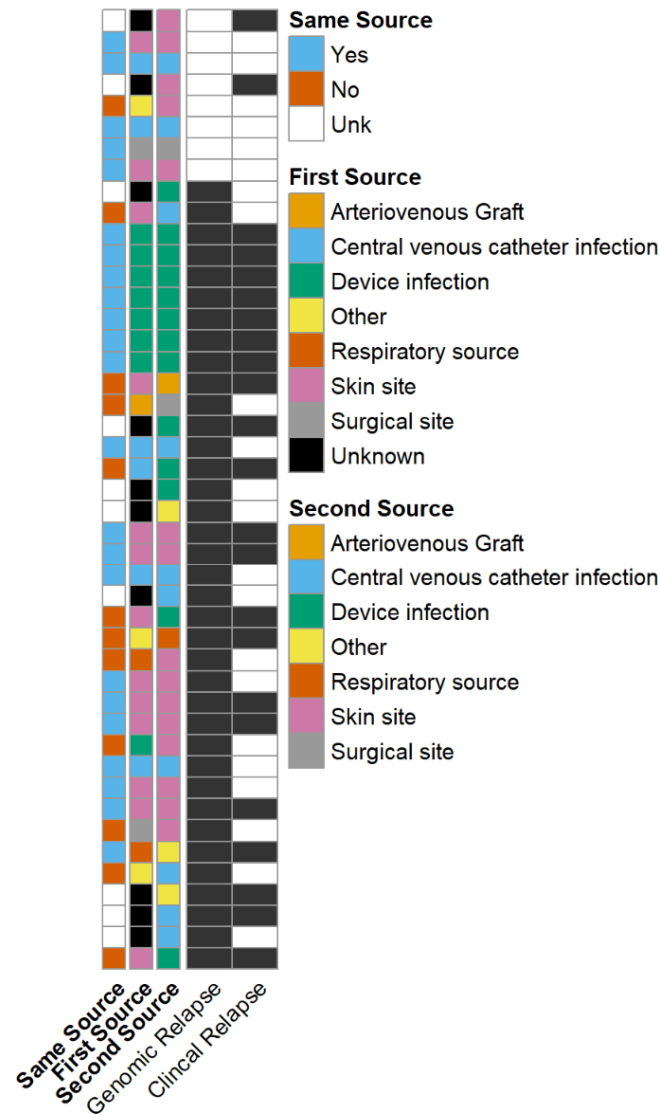
Figure 3: Clinical and genomic definitions of relapse are discordant. Pairs of isolates from all recurrent infections were compared and identified as relapse (filled black rectangle) or new infections (white rectangles) based on a genomic definition or clinical definition, for a total of 45 pairs. The suspected source type of each infection within the episode was identified, with the least recent isolate associated with the "First Source" and the most recent isolate associated with the "Second Source." Clinical source type was additionally assessed to determine if the first and second source were anatomically from the same source.

A

B

C
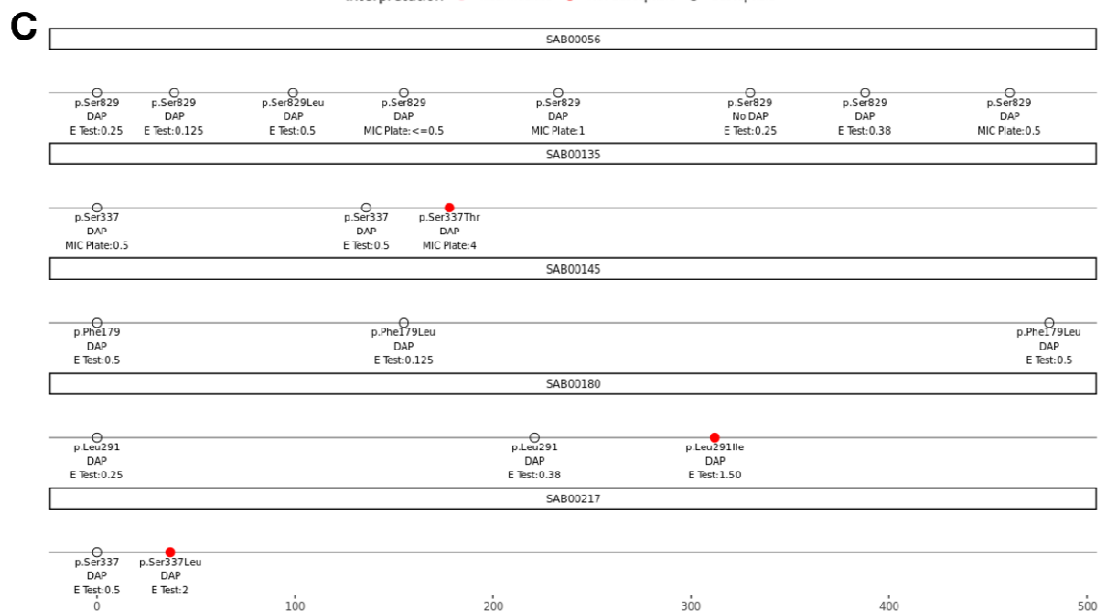
558 **Figure 4: Commonly mutated genes among relapse lineages are associated with antibiotic**

559 **resistance phenotypes.** (A) Unique non-synonymous (NS) 354]mutations in coding regions of

560 the genome were quantified by gene from isolates of relapse-associated infections and

561 summarized relative to the subject from which the isolate was collected. Each unique mutation

562 was annotated with the clonal background of the lineage from which the set of relapses were

563 derived. For lineages with *rpoB* mutations (B) and *mprF* mutations (C), a timeline (days since

564 index infection) was created for each set of relapsing infections by the subject experiencing that

565 set of relapses. Individual episodes were annotated with the amino acid changes detected in the

566 respective genes, the clinical assay used to assess minimum inhibitory concentration (MIC) and

567 the corresponding MIC, and whether the patient was exposed to rifampin (RIF)(B) or

568 daptomycin (DAP)(C). Dots are colored based on the clinical assay determination of drug

569 susceptibility to RIF or DAP.
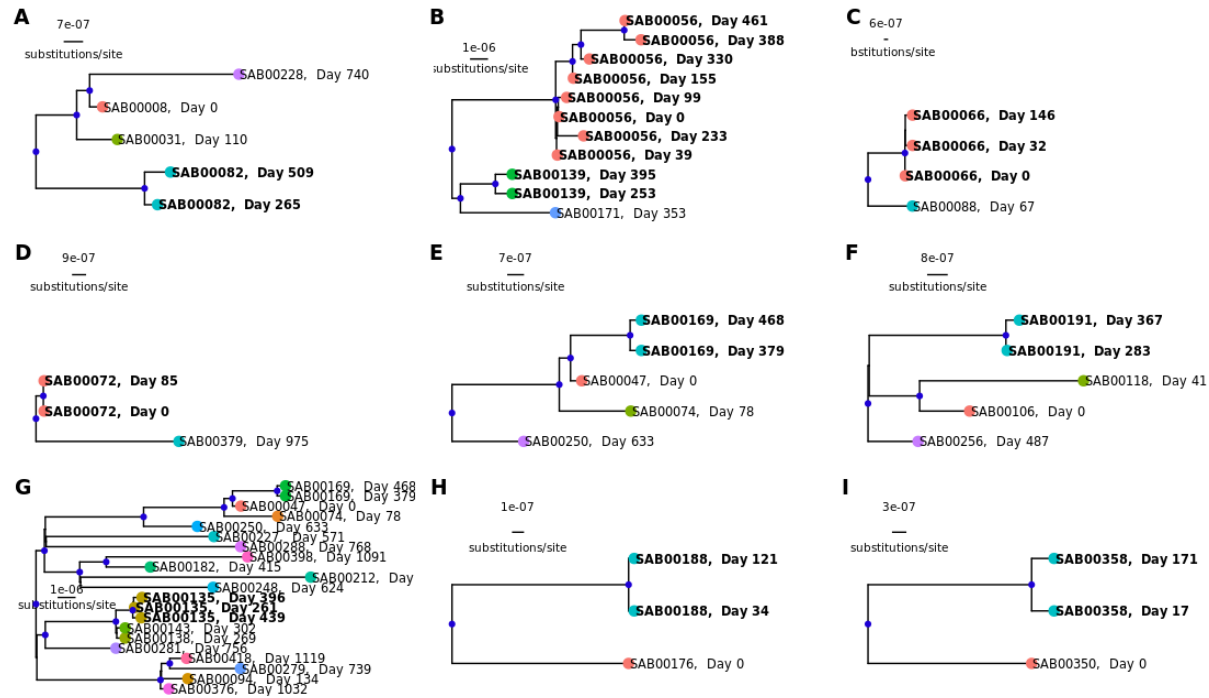
570
571
572
573

Figure 5: Non-relapse associated isolates cluster separately from relapse-associated isolates.

A tree based on core genes was generated using the unequal transition/transversion rate plus
empirical base frequencies model to investigate branching positions. Clusters were investigated
when at least one relapse-associated isolate genome differed by 25 SNPs or fewer from an isolate
from a different subject. Subtrees were extracted based on the most-recent common ancestor
shared by relapse subject isolates and the clustered additional subjects. Nodes denoted with a
blue dot indicate ultrafast bootstrap values and Shimodaira-Hasegawa approximate likelihood
ratio test values that are greater than 70. Bolded tip labels indicate an isolate that is part of a
relapse. Tips are annotated with the patient subject IDs and the number of days at which the
isolate was collected relative to the earliest isolate in the cluster. Branch lengths are scaled in
substitutions per site.

589 **Tables**

590

591 **Table 1. Clinical and demographic characteristics of subjects at the time of the first MRSA**

592 **bacteremia episode for 311 subjects surviving this episode, comparing those with a**

593 **recurrent and those without a recurrent MRSA bacteremia episode**

Patient Attribute	Overall (N = 311)	No recurrent episode (N = 279)	Recurrent episode (N = 32)	p-value
Patient Demographics				
Age at Diagnosis (median, IQR)	54 (41-65)	54 (41-65)	52.5 (33-63)	0.40
Sex				
Male	182 (58.5%)	166 (59.5%)	16 (50%)	0.398c
Female	129 (41%)	113 (40.5%)	16 (50%)	
Race				0.214[c]
Asian	5 (2%)	4 (1%)	1 (3%)	
Black	131 (42%)	121 (43%)	10 (31%)	
White	146 (46%)	128 (46%)	18 (56%)	

More than one race	2 (<1%)	1 (<1%)	1 (3%)	
Other Race	9 (3%)	8 (3%)	1 (3%)	
Unknown	18 (6%)	17 (6%)	1 (3 %)	
Ethnicity				0.644[c]
Hispanic or Latino	14 (4.5%)	12 (4%)	2 (6%)	
Not Hispanic or Latino	297 (95%)	267 (96%)	30 (94%)	
Patient Comorbid Conditions				
Chronic skin disease	32 (10%)	29 (10%)	3 (9%)	1[c]
Diabetes mellitus	105 (46%)	92 (33%)	13 (41%)	0.503[b]
Cancer	40 (13%)	37 (13%)	3 (9%)	0.8804[c]
Respiratory disease	68 (22%)	59 (21%)	9 (28%)	0.5796[b]
Cardiovascular disease	148 (47.5%)	125 (45%)	19 (59%)	0.168[b]
Liver disease	30 (10%)	26 (9%)	4 (12.5%)	0.5813[c]

Kidney disease	83 (27%)	72 (26%)	11 (34%)	0.4083[b]
Hemodialysis in the last 12 months?	59 (19%)	51 (18%)	8 (25%)	0.3471[c]
Current intravenous drug use	74 (24%)	67 (24%)	7 (22%)	1[c]
Complication of bacteremia				
Infective endocarditis	50 (16%)	46 (16.5%)	4 (12.5%)	0.82
Epidural abscesses	18 (6%)	15 (5%)	3 (9%)	0.4734[c]
Soft tissue abscess	41 (13%)	36 (13%)	5 (16%)	0.6333[c]
Osteomyelitis	67 (21.5%)	60 (21.5%)	7 (22%)	1[c]
Septic pulmonary emboli	41 (13%)	40 (14%)	1 (3%)	0.189[c]
Other metastatic complication(s)	43 (14%)	40 (14%)	3 (9%)	0.593[c]
Involvement of foreign body in the	69 (22%)	62 (22%)	7 (22%)	1[c]

bacteremia				
Healthcare acquisition type of index infection				0.1373[c]
Community-associated	38 (12%)	36 (13%)	2 (6%)	
Healthcare-associated	61 (20%)	58 (21%)	3 (9%)	
Healthcare-associated-Community onset	212 (68%)	185 (66%)	27 (84%)	
How many antibiotics was the patient administered during the index MRSAB?	2 (1-2)	2 (1-2)	2 (1-3)	1[a]
Vancomycin	295 (95%)	264 (95%)	31 (97%)	1[c]
Vancomycin	28 (10.5-42)	28 (11-42)	28 (6 – 42)	0.569[a]

duration (median days, IQR))[d]				
Daptomycin	123 (40%)	105 (38%)	18 (56%)	0.0644[b]
Daptomycin duration (median days, IQR)[d]	33 (12-42)	32.5 (12-42)	38 (21-42.5)	0.433[a]

594   a Kruskal-Wallis test

595   b Chi-square test

596   c Fisher Exact test

597   d Among subjects with exposure

598

599

600

601

602

603

604

605

606

607

608

609

610

611 **Table 2: Neutrality index calculations of the whole genome and bacteremia-associated**

612 **genes for all MRSA bacteremia isolates and for relapse-associated lineages**

Group	Fixed, Non-synonymous	Fixed, Synonymous	Polymorphic, Non-synonymous	Polymorphic, Synonymous	McDonald-Kreitman Value	P-value (G-test)
All isolates - whole genome	17271	24529	1381	1812	0.92	**0.03**
All isolates - bacteremia genes	232	330	19	22	0.81	0.53
Relapse - whole genome	2689	5326	45	12	0.133	**< 0.01**
Relapse - bacteremia genes	28	81	5	0	–	–

613