

Sussex Research

Antimicrobial resistance determinants are associated with *Staphylococcus aureus* bacteraemia and adaptation to the healthcare environment: a bacterial genome-wide association study

Bernadette C Young, Chieh-Hsi Wu, Jane Charlesworth, Sarah Earle, James R Price, N Claire Gordon, Kevin Cole, Laura Dunn, Elian Liu, Sarah Oakley, Heather Godwin, Rowena Fung, Ruth Miller, Kyle Knox, Antonina Votintseva, T Phuong Quana, Robert Tilley, Matthew Scarborough, Derrick W Crook, Timothy E Peto, A Sarah Walker, Martin Llewelyn, Daniel J Wilson

Publication date

27-11-2021

Licence

This work is made available under the [CC BY 4.0](#) licence and should only be used in accordance with that licence. For more information on the specific terms, consult the repository record for this item.

Document Version

Published version

Citation for this work (American Psychological Association 7th edition)

Young, B. C., Wu, C.-H., Charlesworth, J., Earle, S., Price, J. R., Gordon, N. C., Cole, K., Dunn, L., Liu, E., Oakley, S., Godwin, H., Fung, R., Miller, R., Knox, K., Votintseva, A., Quana, T. P., Tilley, R., Scarborough, M., Crook, D. W., ... Wilson, D. J. (2021). *Antimicrobial resistance determinants are associated with Staphylococcus aureus bacteraemia and adaptation to the healthcare environment: a bacterial genome-wide association study* (Version 1). University of Sussex. <https://hdl.handle.net/10779/uos.23484737.v1>

Published in

Microbial Genomics

Link to external publisher version

<https://doi.org/10.1099/mgen.0.000700>

Copyright and reuse:

This work was downloaded from Sussex Research Open (SRO). This document is made available in line with publisher policy and may differ from the published version. Please cite the published version where possible. Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners unless otherwise stated. For more information on this work, SRO or to report an issue, you can contact the repository administrators at sro@sussex.ac.uk. Discover more of the University's research at <https://sussex.figshare.com/>

Antimicrobial resistance determinants are associated with *Staphylococcus aureus* bacteraemia and adaptation to the healthcare environment: a bacterial genome-wide association study

Supplemental information

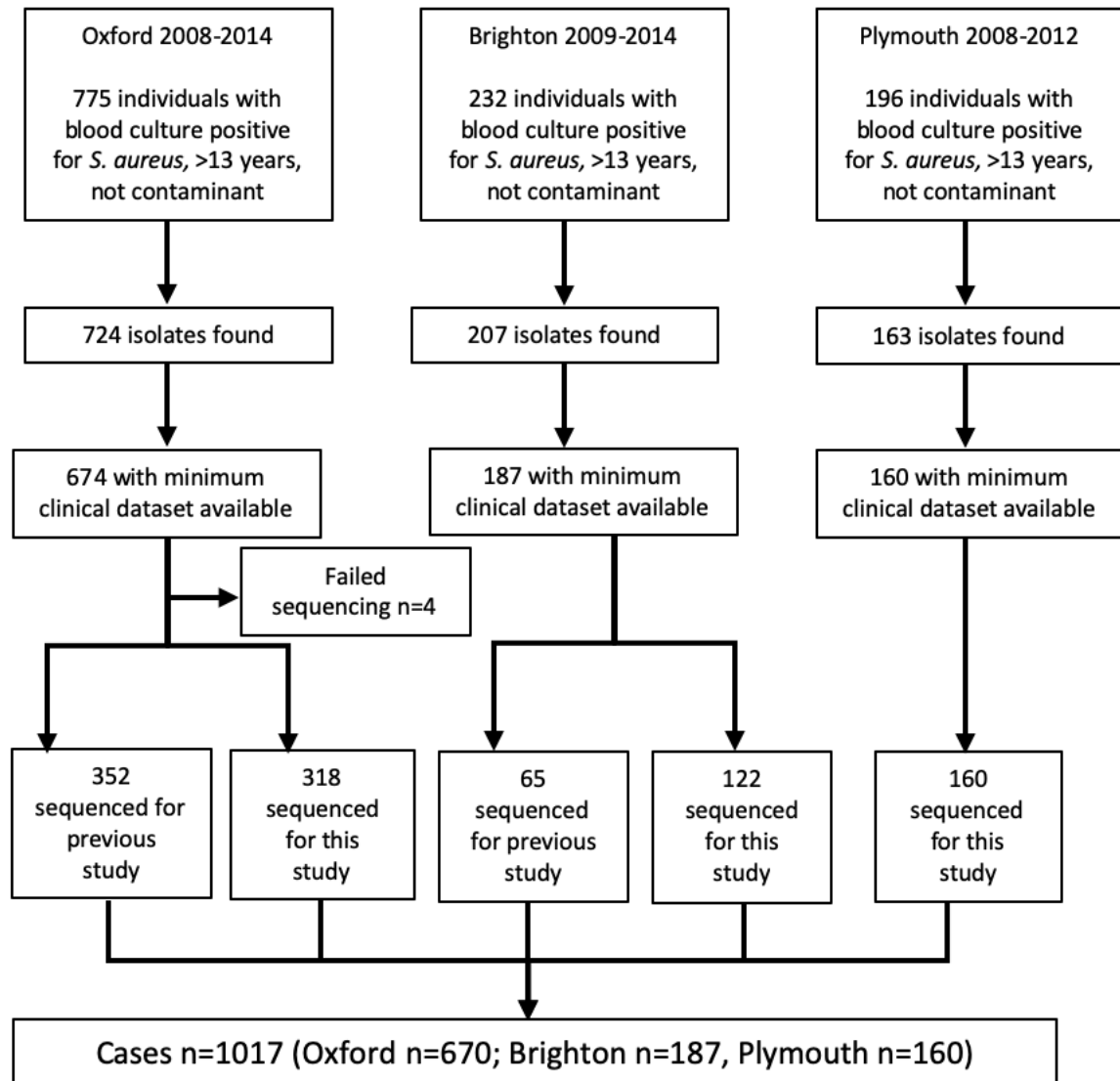


Figure S1: Flow diagram of cases screened for inclusion in the study

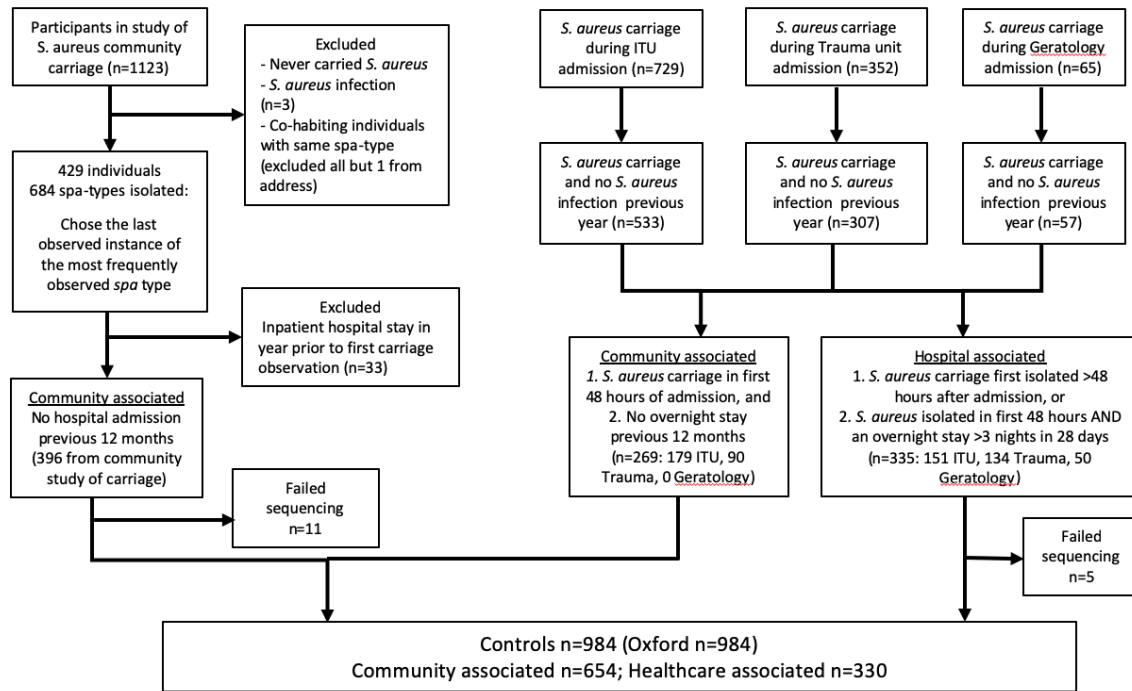


Figure S2: Flow diagram of controls screened for inclusion in the study

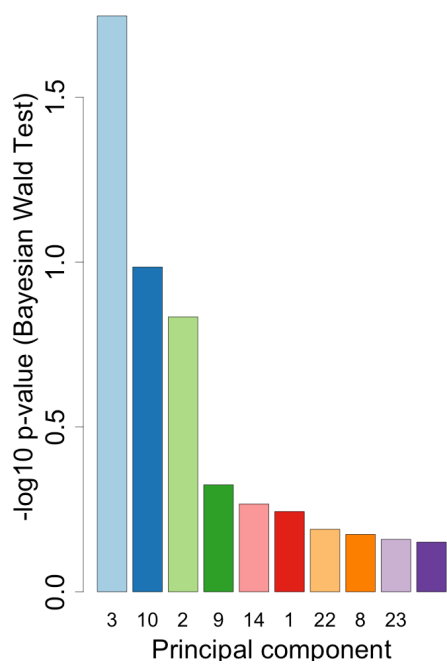
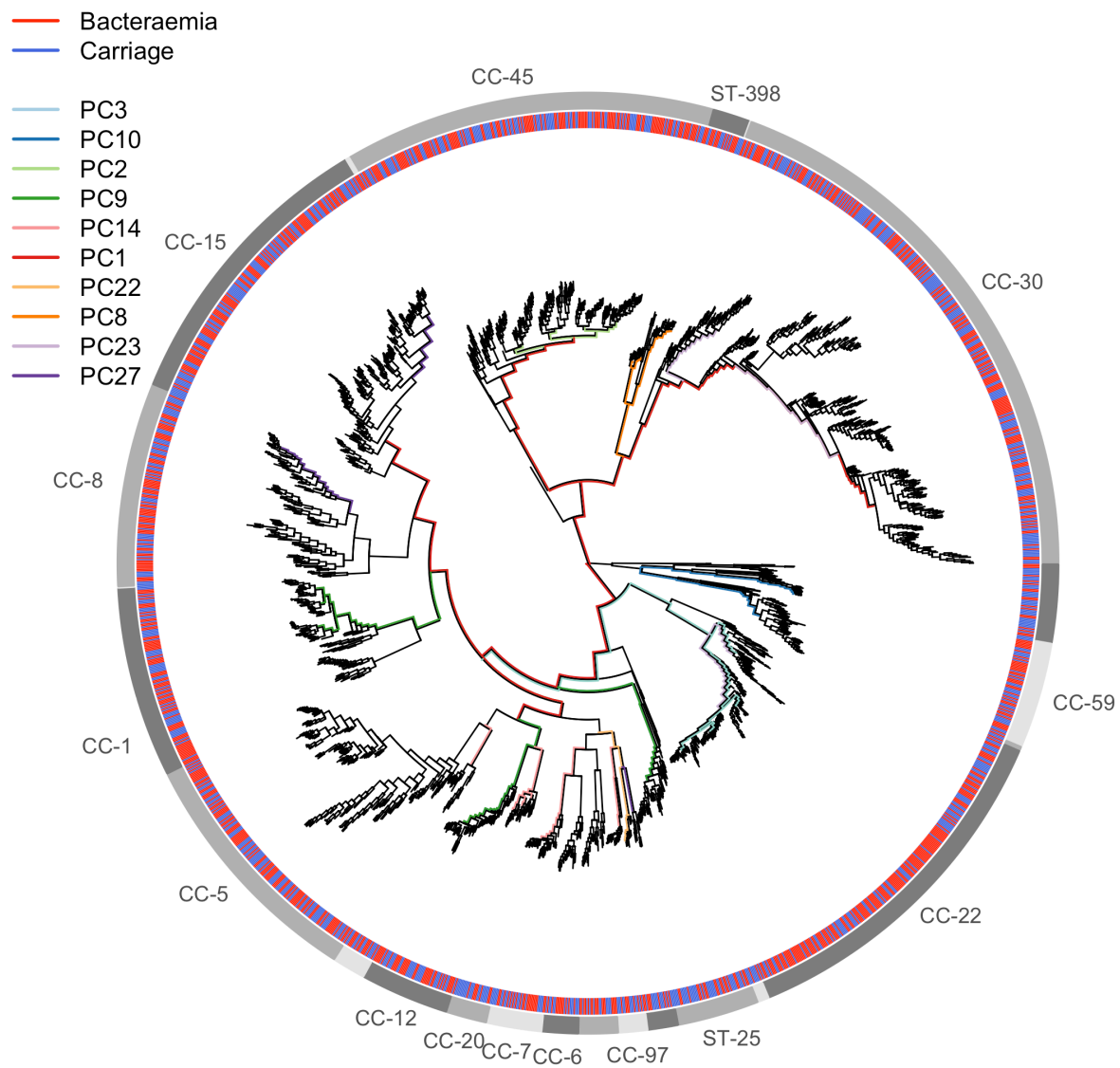


Figure S3: Association of Principal Components with bacteraemia-vs-carriage.

(A) The branches corresponding to the 10 lineages most significantly associated with bacteraemia-vs-carriage are coloured on a maximum likelihood phylogeny of the study isolates. Branch lengths have been square-root transformed to discriminate closely related lineages. Clonal complexes are denoted in the outermost ring. The second outermost ring indicates isolate source (blue carriage, red bacteraemia) **(B)** Significance (negative \log_{10} p -values) of the 10 most significantly associated PCs. A Bonferroni corrected threshold for significance is $10^{-4.6}$.

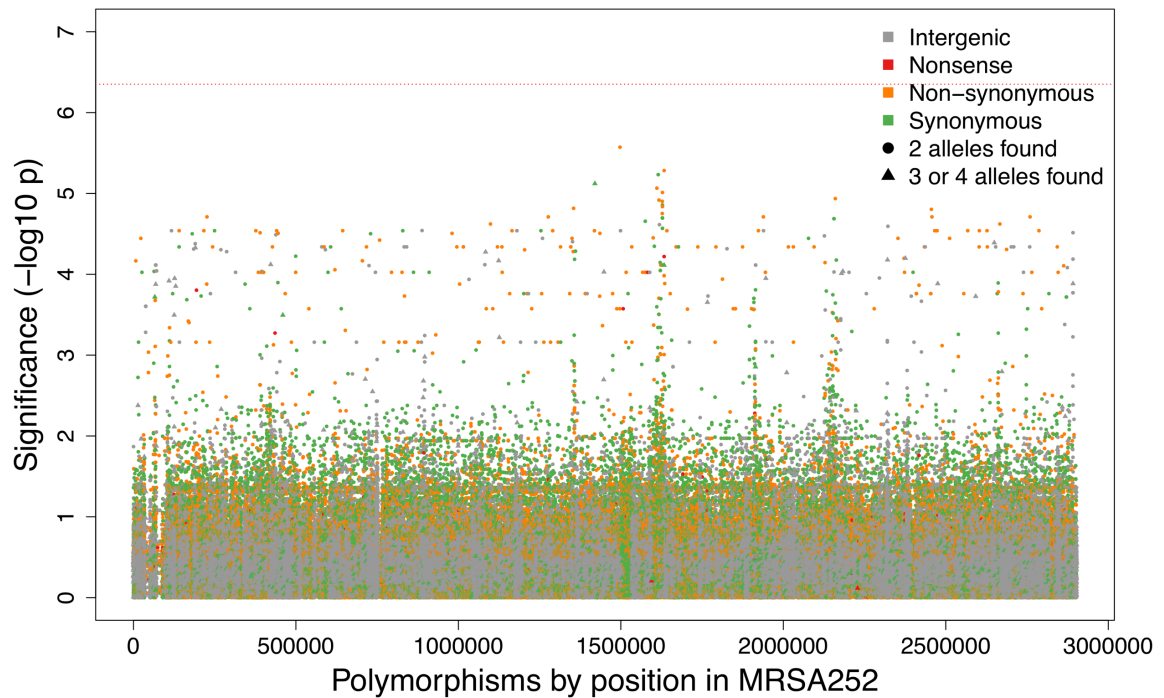


Figure S4: Manhattan plot of SNP associations with bacteraemia-vs-carriage, controlling for population structure and HA or CA origin. The significance of each SNP ($-\log_{10} p$ value LRT) is plotted according to location on the 2.9MB MRSA252 reference genome, with control for population structure and HA-vs-CA origin. SNPs are coloured according to their predicted effect on protein, and shaped according to the number of alleles found at that site in the study set. A Bonferroni-corrected significance threshold based on the number of phylopatterns ($n=112,292$) is plotted in red ($10^{-6.4}$), assuming a family-wise error rate of 5%.

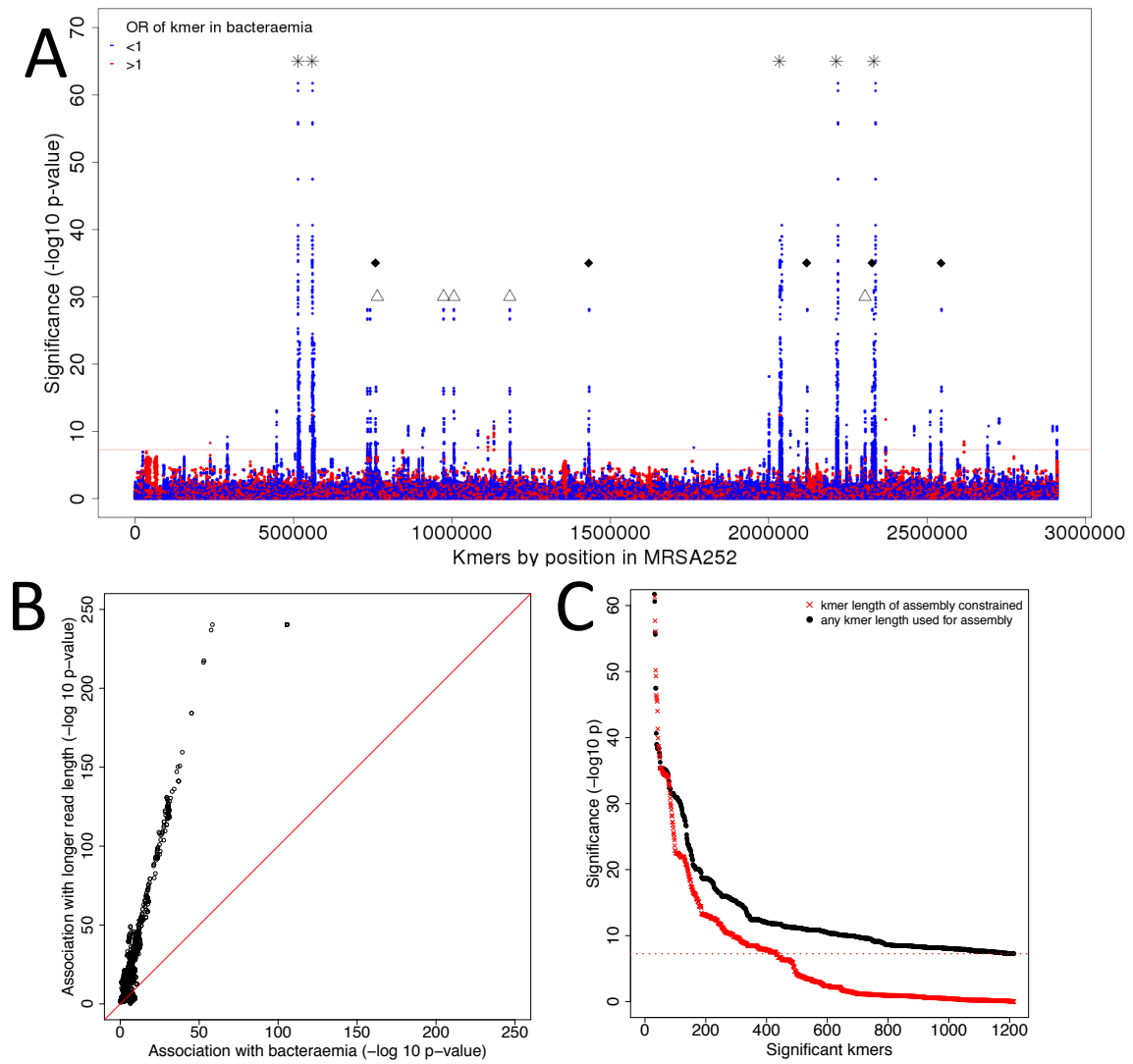


Figure S5: Confounding of kmer study by sequencing read length. (A) Manhattan plot of kmers associated with phenotype in study of 2001 isolates. Kmers are mapped to reference genome and plotted against statistical significance of association with case-control-status. Kmers found more frequently in carriage are mapped in blue, kmers found more frequently in bacteraemia are mapped in red. A Bonferroni corrected threshold for significance is plotted in red ($10^{-7.3}$). The positions of several repeat regions are marked: 16S rDNA (*), ISS (Δ) and ISSSau3 (\blacklozenge). **(B)** The kmers most associated with bacteraemia-vs-carriage were confounded with sequence length. We plotted significance of association with both bacteraemia-vs-carriage (X axis) and sequencing read length (Y axis), using a χ^2 test, without correction for population structure. **(C)** Kmer detection and association testing with GEMMA was repeated, this time on new assemblies, built using Velvet with a maximum kmer length of 79bp. The signal of the most significant kmers based on the original assemblies (black) dropped to below statistical significance when the assemblies were built with a constrained kmer length (red).

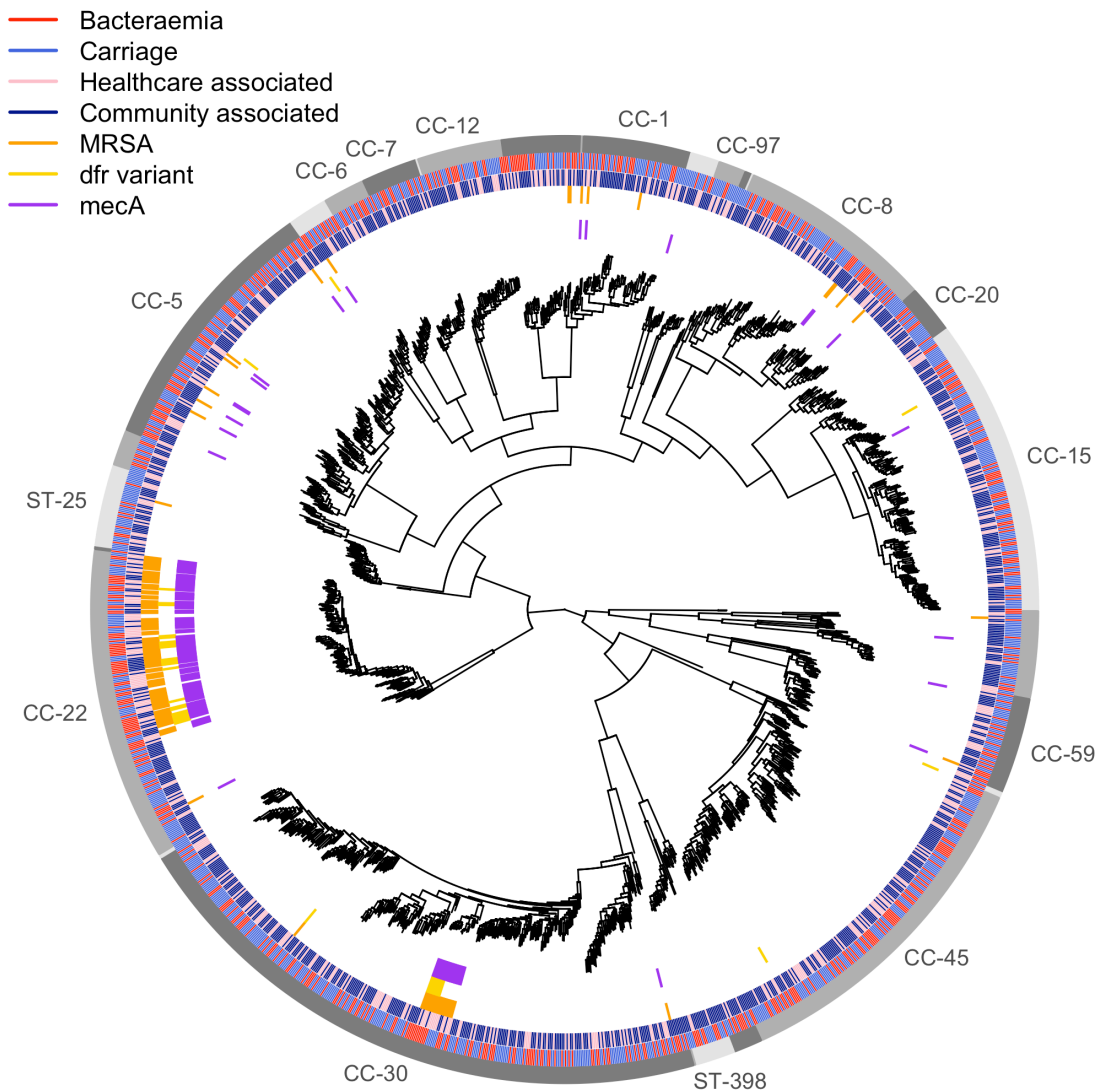
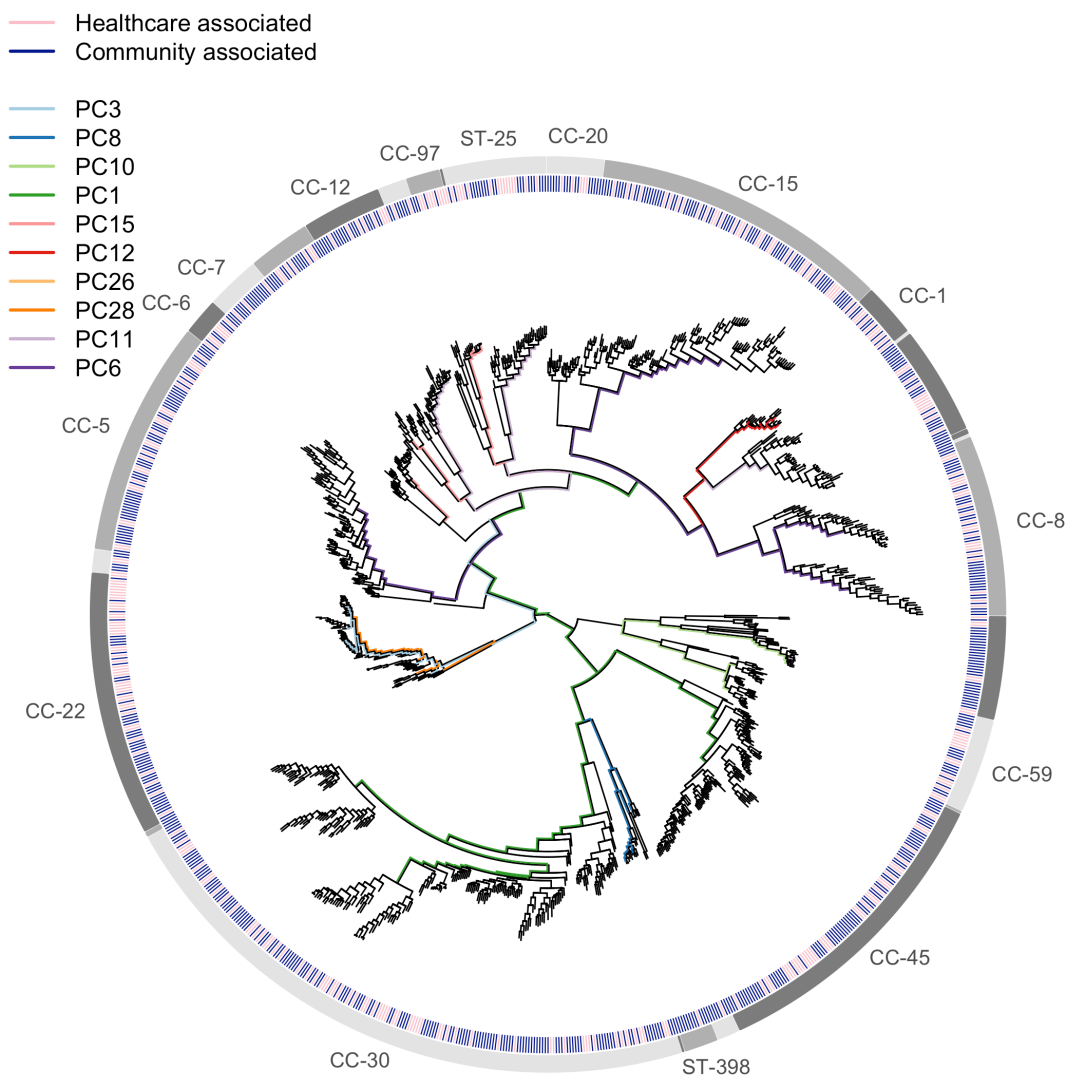


Figure S6 Presence of kmers that are significantly associated with bacteraemia-vs-carriage among 1610 isolates.

Branch lengths have been square-root transformed to better discriminate closely related lineages. The outer ring indicates clusters with a shared lineage named by the clonal complex (or ST if only a single ST was in the cluster). The second outermost ring indicates isolate source (blue carriage, red bacteraemia). The third outermost ring indicates whether each isolate was community or healthcare associated. The fourth ring indicates whether an isolate was phenotypically methicillin resistant (orange). The fifth ring marks the presence of kmers that encode a trimethoprim conferring variant in *dfrB* (yellow). The innermost ring indicates the presence of kmers mapping to the sequence of *mecA* (SAR0039).

A



B

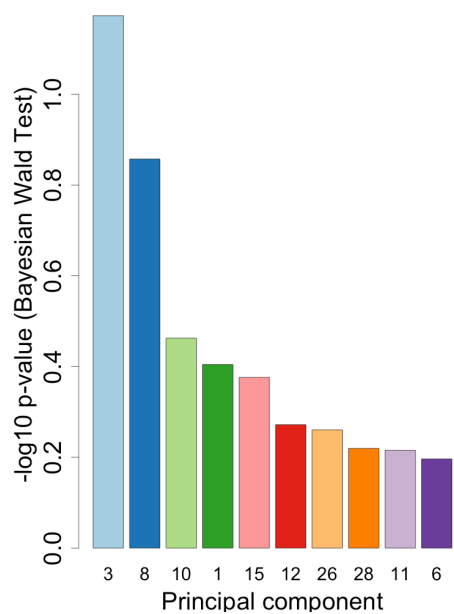


Figure S7: Association of Principal Components with HA-vs-CA carriage. (A) The branches corresponding to the 10 lineages most significantly associated with HA-vs-CA carriage are coloured on a maximum likelihood phylogeny of the study isolates. Branch lengths have been square-root transformed to discriminate closely related lineages. Clonal complexes are denoted in the outermost ring. The second outermost ring indicates isolate source (blue community, pink healthcare) **(B)** Significance (negative $\log_{10} p$ -values) of the 10 most significantly associated PCs. A Bonferroni corrected threshold for significance is $10^{-4.3}$.

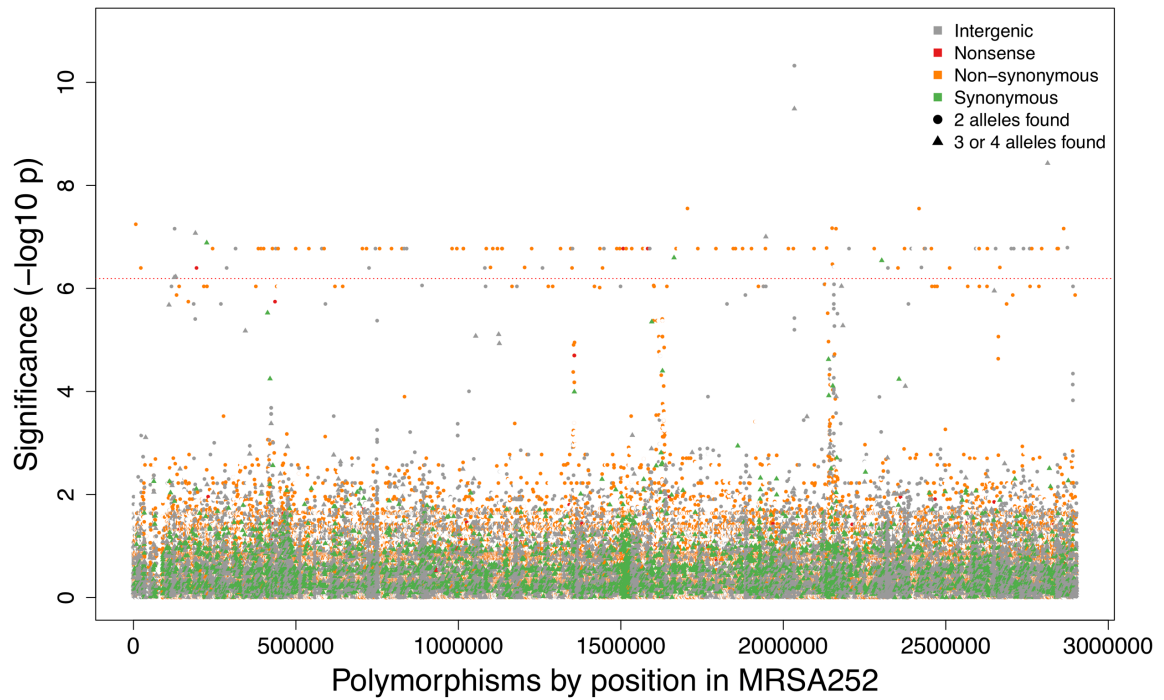


Figure S8: Manhattan plot of SNP associations with HA-vs-CA *S. aureus* carriage. The significance of each SNP ($-\log_{10} p$ value) is plotted according to location on the 2.9MB MRSA252 reference genome, with control for population structure. SNPs are coloured according to their predicted effect on protein, and shaped according to the number of alleles found at that site in the study set. We applied Bonferroni correction to the intended 5% false positive rate given the number of SNP phylopatterns to obtain a genome-wide significance threshold of $10^{-6.2}$ (red dotted line).

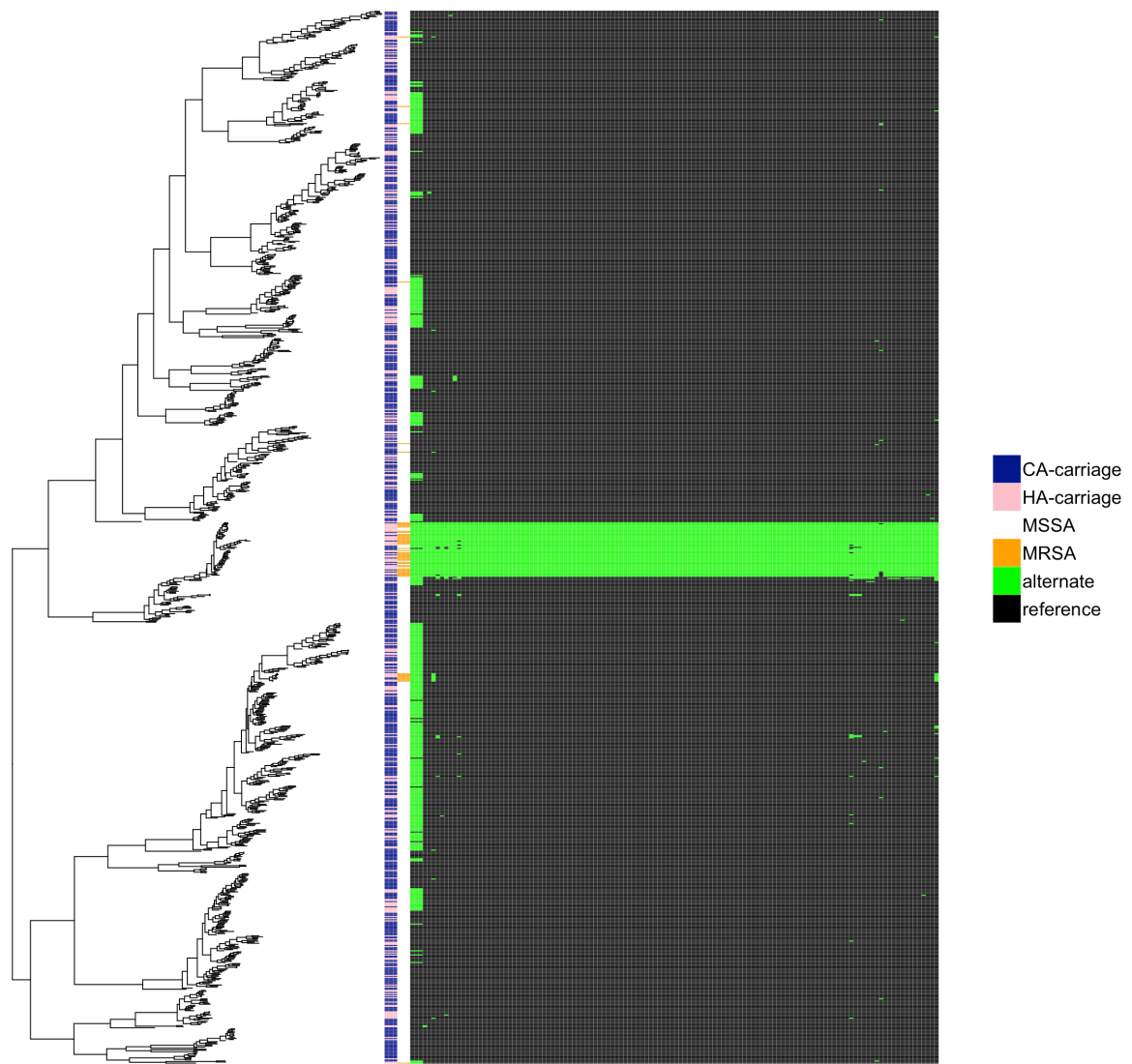


Figure S9: Predicted presence or absence of SNPs significantly associated with HA-vs-CA carriage among the carriage population. A maximum likelihood phylogeny of 984 carriage isolates is plotted with a matrix representing phenotypes concatenated across the most significant SNP genotypes, in order of the significance of association. A maximum likelihood tree of carriage isolates is annotated with the predicted origin of carriage (pink HA, blue CA), the presence of MRSA (orange). The remaining columns represent the presence of the reference (black) or alternate (green) allele for all significant SNPS (after imputation) which were biallelic in the study population.

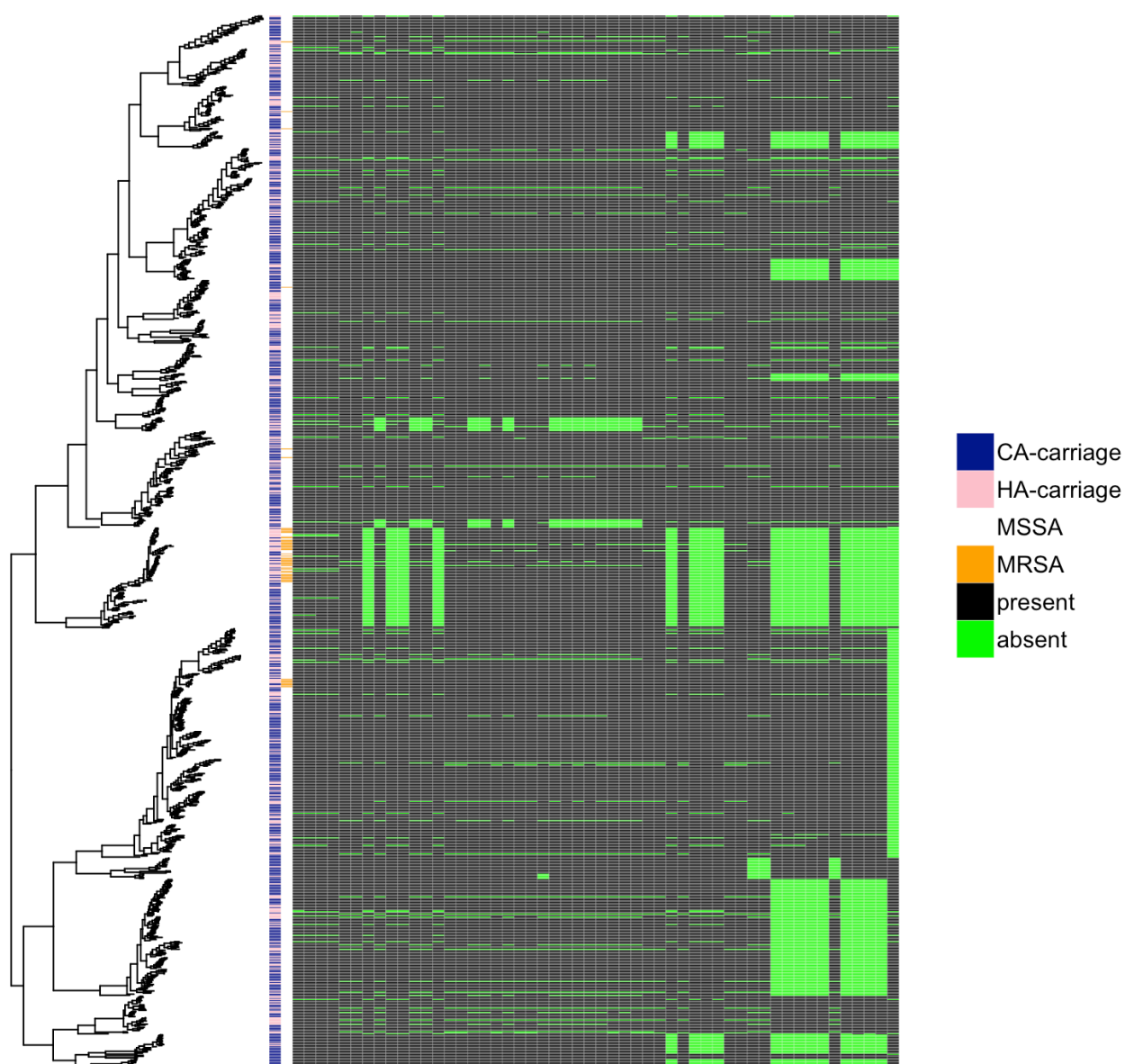


Figure S10: Presence of significant kmers in the population of CA and HA carriage. A maximum likelihood tree of carriage isolates is annotated with the predicted origin of carriage (pink HA, blue CA), the presence of MRSA (orange) and the presence (black) or absence (green) of each pattern of kmers mapping to *prsA*, in decreasing order of statistical significance left to right.

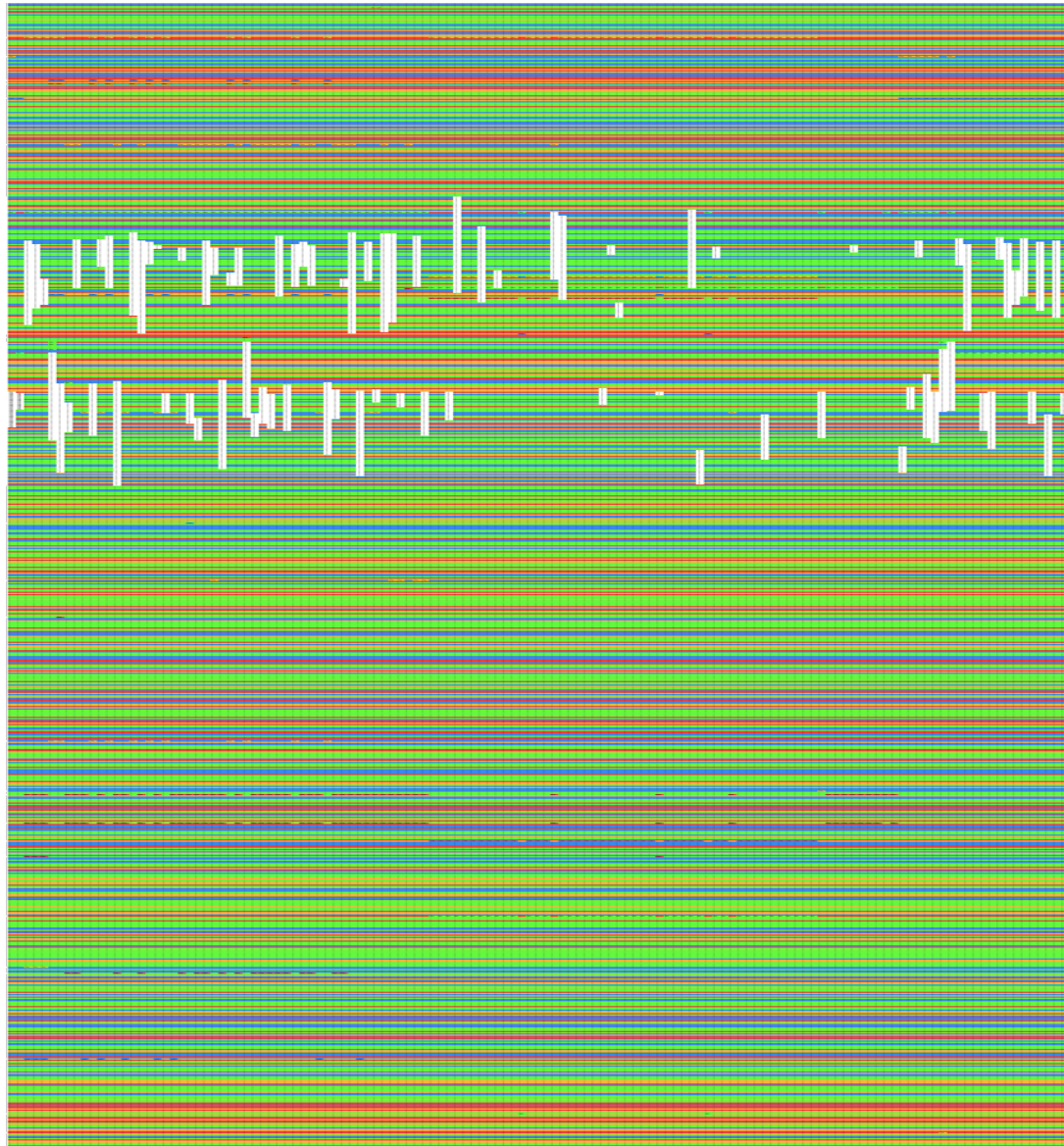


Figure S11: Predicted sequence of *prsA* in 131 HA-carriage genomes. The predicted sequences from 131 HA carriage genomes with the lowest number of significant *prsA* kmers found were aligned against the reference (NC_002952.2:2011797-2012759), and are displayed here aligned vertically and coloured by nucleotide (A green, C orange, G red, T blue, null – white) A plurality of variation is found, including SNPs and short deletions up to 87bp. Deletions are found across bases 164-406; the absence of kmers mapping to bases 177-394 was associated with HA carriage (Figure 3B).

| Focus | n | % | UKCIRG (%) |
|----------------------------|----------|----------|-------------------|
| Central venous catheter | 101 | 10.9 | 18.2 |
| Peripheral venous catheter | 47 | 5.1 | 5.8 |
| Endocarditis | 42 | 4.7 | 5.8 |
| Soft tissue | 244 | 26.4 | 19.5 |
| Respiratory | 52 | 5.6 | 3.7 |
| Urinary | 35 | 3.8 | (NA) |
| Thrombus | 11 | 1.2 | (NA) |
| Vascular implant | 17 | 1.8 | (NA) |
| Other | 134 | 14.5 | 13.6 |
| Not established | 235 | 25.5 | 18.8 |
| (Missing data) | 94 | 9.2% | |

Table S1: Focus of infection for cases (n=1017) of *Staphylococcus aureus* bacteraemia included, and compared with focus of infection reported in a large cohort of SAB in the UK.⁶

| Centre | All cause mortality by 7 days | All cause mortality by 30 days | All cause mortality by 90 days | Mortality data missing |
|----------------|-------------------------------|--------------------------------|--------------------------------|------------------------|
| Oxford | 94 | 180 | 228 | 6/670 (0.9) |
| N=664 | (14.1) | (27.1) | (34.3) | |
| Brighton | 15 | 27 | 38 | 70/187 (37.4) |
| N=117 | (12.8) | (23.1) | (32.5) | |
| Plymouth | 14 | 38 | 54 | 6/160 (3.7) |
| N=154 | (9.1) | (24.7) | (35.1) | |
| All | 123 | 245 | 320 | 92/1017 (9.0) |
| N = 925 | (13.3) | (26.5) | (34.6) | |

Table S2: Mortality at 7, 30 and 90 days, total (percentage). Differences between centres and 7, 30 and 90 days were not statistically significant ($p=0.25$, $p=0.6$ and $p=0.9$, χ^2 test with 2 degrees of freedom)

Table S3: SNPs showing significant association with HA-vs-CA carriage

Table S4: Kmers showing significant association with HA-vs-CA carriage

Table S5: Summary of data availability for all strains included in study, including data deposition in SRA