All Staphylococcus aureus bacteraemia-inducing strains can cause infective endocarditis

Results of GWAS and experimental animal studies

Bastien, Sylvère; Meyers, Severien; Salgado-Pabón, Wilmara; Giulieri, Stefano G.; Rasigade, Jean-Phillipe; Liesenborghs, Laurens; Kinney, Kyle J.; Couzon, Florence; Martins-Simoes, Patricia; Moing, Vincent Le; Duval, Xavier; Holmes, Natasha E.; Bruun, Niels Eske; Skov, Robert; Howden, Benjamin P.; Fowler Jr., Vance G.; Verhamme, Peter; Andersen, Paal Skytt; Bouchiat, Coralie; Moreau, Karen; Vandenesch, François

Published in:
Journal of Infection

DOI (link to publication from Publisher):
10.1016/j.jinf.2022.12.028

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Journal of Infection 86 (2023) 123–133

Contents lists available at ScienceDirect

Journal of Infection

journal homepage: www.elsevier.com/locate/jinf

All *Staphylococcus aureus* bacteraemia-inducing strains can cause infective endocarditis: Results of GWAS and experimental animal studies

Sylvère Bastien[a], Severien Meyers[b,1], Wilmara Salgado-Pabón[c,1], Stefano G. Giulieri[d,e], Jean-Phillipe Rasigade[a,f], Laurens Liesenborghs[b], Kyle J. Kinney[g], Florence Couzon[a], Patricia Martins-Simoes[a,f], Vincent Le Moing[h], Xavier Duval[i,j], Natasha E Holmes[e], Niels Eske Bruun[k,l], Robert Skov[m], Benjamin P Howden[d,e], Vance G. Fowler Jr.[n,o], Peter Verhamme[b], Paal Skytt Andersen[m], Coralie Bouchiat[a,f,2], Karen Moreau[a,2], François Vandenesch[a,f,2,*]

[a] *CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France*
[b] *Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, University of Leuven, Leuven, Belgium*
[c] *Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, USA*
[d] *Department of Microbiology and Immunology and Department of Infectious Diseases, The University of Melbourne at the Doherty Institute for Infection and Immunity; Victorian Infectious Disease Service, The Royal Melbourne Hospital, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia*
[e] *Department of Infectious Diseases, Austin Health, Heidelberg, Australia*
[f] *Centre National de Référence des Staphylocoques, Institut des Agents Infectieux, Hospices Civils de Lyon, F-69004, Lyon, France*
[g] *Department of Microbiology and Immunology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA, USA*
[h] *Service de Maladies Infectieuses, CHU de Montpellier, France*
[i] *Hôpital Bichat Claude Bernard, AP-HP, Paris, France*
[j] *Inserm CIC 1425, Inserm UMR-1137 IAME, Cité Paris University, UFR de Médecine-Bichat, Paris, France*
[k] *Clinical Institute, Copenhagen and Aalborg University, Aalborg, Denmark*
[l] *Department of Cardiology, Zealand University Hospital Roskilde, Roskilde, Zealand, Denmark*
[m] *Department of Bacteria, Parasites, and Fungi, Statens Serum Institut, Copenhagen, Denmark*
[n] *Duke University Medical Center, Durham, NC USA*
[o] *Duke Clinical Research Institute, Durham, NC USA*

ARTICLE INFO

Article history:
Accepted 24 December 2022
Available online 2 January 2023

Keyword:
Infective endocarditis
Bacteraemia
Genome-wide association study
Experimental animal model
Staphylococcus aureus

SUMMARY

Objectives: We aimed at determining whether specific *S. aureus* strains cause infective endocarditis (IE) in the course of *Staphylococcus aureus* bacteraemia (SAB).
Methods: A genome-wide association study (GWAS) including 924 *S. aureus* genomes from IE (274) and non-IE (650) SAB patients from international cohorts was conducted, and a subset of strains was tested with two experimental animal models of IE, one investigating the early step of bacterial adhesion to inflamed mice valves, the second evaluating the local and systemic developmental process of IE on mechanically-damaged rabbit valves.
Results: The genetic profile of *S. aureus* IE and non-IE SAB strains did not differ when considering single nucleotide polymorphisms, coding sequences, and k-mers analysed in GWAS. In the murine inflammation-induced IE model, no difference was observed between IE and non-IE SAB strains both in terms of adhesion to the cardiac valves and in the propensity to cause IE; in the mechanical IE-induced rabbit model, there was no difference between IE and non-IE SAB strains regarding the vegetation size and CFU.

* Corresponding author.
E-mail address: francois.vandenesch@univ-lyon1.fr (F. Vandenesch).
[1] equal authorship
[2] equal senior authorship

https://doi.org/10.1016/j.jinf.2022.12.028

Conclusion: All strains of *S. aureus* isolated from SAB patients must be considered as capable of causing this common and lethal infection once they have accessed the bloodstream.

Introduction

Staphylococcus aureus bacteraemia (SAB) is a severe condition whose incidence ranges from 38.2 to 45.7 per 100,000 person-years in the US.[1,2] In 10 to 25% of SAB cases, the infection localises to the valves or endocardial surfaces, leading to infective endocarditis (IE)[3,4] and thereby to enhanced morbidity and mortality. To exclude IE upon diagnosing SAB, the use of either trans-thoracic or transoesophageal echocardiography is recommended[5,6] but is far from being systematic: a recent study analysing 668,423 hospitalizations for SAB from the US National Inpatient Sample database (2001–2014) reported that only 11% patients in 2001 to 15% in 2014 had echocardiogram.[7] In this context, score-based prediction rules such as the VIRSTA, PREDICT, or POSITIVE scores have been proposed to quantify the risk of IE in patients with a SAB diagnosis and guide the use of echocardiography.[8-10] However, none of these scores take into account the possibility that certain strains of *S. aureus* are more likely than others to cause IE. Previous studies have compared clinical IE strains with other disease isolates using molecular methods in order to point out which *S. aureus* shared genetic background/lineage and virulence factors are actually crucial in causing endocarditis; however, such studies have yielded contradictory results.[11-14] Thus the lineages (CC=clonal complexes) CC12 and CC20 were associated with IE isolates and not with non-IE SAB and skin and soft tissue infection (SSTI) isolates.[11] Similarly, CC30 and several adhesins and enterotoxins were associated with IE but not with SSTI, nasal carriage, and non-IE SAB isolates.[12,15] The invasive isolates responsible for non-IE SAB were found to be less cytotoxic than non-invasive or colonising isolates, the differences between diseases were supported by polymorphisms at several loci.[16] However, when comparing IE with non-IE bloodstream isolates of *S. aureus*, weak[17] or no significant differences[18] between strains were observed. Similarly, there was no significant difference in the phenotypic traits known or hypothesised to be involved in IE between IE versus non-IE SAB strains.[17] More recently, a genome-wide association study (GWAS) performed on 241 strains (120 definite IE and 121 non-IE SAB) using a variety of bioinformatics approaches including single nucleotide polymorphism (SNP) analysis, accessory genome analysis, and k-mers based analysis did not identify clear *S. aureus* genetic markers associated with the occurrence of IE in the course of bacteraemia.[19] However, the latter study had some limitations, including a small population size and limited geographical distribution, emphasising the need to conduct GWAS on a larger sample size.

In order to provide a more robust assessment of the potential role of *S. aureus* factors in the occurrence of IE during bacteraemia, we performed a GWAS analysis of 924 *S. aureus* genomes from SAB patients with or without IE, and completed this genomic approach with two experimental models of IE exploring both early and late events of aortic valve infection.

Methods

Bacterial strains

The strain collection was designed to represent the western and oceanic genetic diversity of *S. aureus* bloodstream isolates isolated from both IE and uncomplicated bacteraemia cases (non-IE SAB) comprising both excluded and possible IE. To this end, several patient cohorts in which all patients underwent trans thoracic (TTE) or trans oesophageal echography (TEE), were combined (Table 1), achieving the number of 274 definite IE according to the modified Duke criteria[20] and 650 non-IE SAB patients as follows: 130 strains (72 IE and 58 non-IE SAB cases) from the French national prospective multicentre cohort VIRSTA,[3] 26 strains (13 IE and 13 non-IE SAB) from the Danish National Staphylococcus aureus Bacteraemia Repository (Statens Serum Institut), 40 strains (20 IE and 20 non-IE SAB) obtained from the *Staphylococcus aureus* Bacteraemia Group (SABG) biorepository from the Duke University,[21] 487 strains (49 IE and 438 non-IE SAB) from two Australasian cohorts of SAB, namely the vancomycin sub-study of the Australian and New Zealand Cooperative on Outcome of Staphylococcal Sepsis (ANZCOSS)[22] and the Vancomycin Efficacy in Staphylococcal Sepsis in Australasia (VANESSA) study,[23] and 241 strains (120 IE and 121 non-IE SAB) collected during two prospective studies in Denmark already analysed by GWAS.[19,24] The AUS/NZ cohort which collected all SAB strains is the only undirected one; its 10% prevalence of IE reflects what could be expected in SAB in general. For the other cohorts, the ratios of IE to non-IE SAB were voluntarily adjusted to obtain approximately 1 IE for 1 non-IE SAB in order to increase the statistical power when using a GWAS. Finally only CC30 were included from *Staphylococcus aureus* Bacteraemia Group (SABG) biorepository from the Duke University, in order to enrich the whole collection in this lineage which is known to be prevalent in SAB in the US.[12]

All patients underwent either TTE or TEE and patients with intracardiac devices were excluded. For animal experiments using batch inoculum, a subset of 12 IE and 12 non-IE SAB CC5 strains, as well as 10 IE and 9 non-IE SAB CC45 strains were selected within the collection of 924 strains to account for the genomic diversity of the CC5 and CC45 lineages. For single strain experiments, four CC5 (two IE and two non-IE SAB) or eight CC5 (four IE and four non-IE SAB) lineages were arbitrarily chosen for the mouse and rabbit models, respectively (see Supplementary excel File).

Bacterial whole genome sequencing

Libraries were prepared using Nextera® or Truseq® DNA (Illumina, San Diego, CA, United States). Bacterial whole genome sequencing was conducted, depending on centres (Table 1), on MiSeq® or NextSeq® or HiSEQ-1500® (Illumina, San Diego, CA, United States) platforms with a read length of 2 × 300 bp, 2 × 250 bp, or 2 × 150 bp. Further details are available in the Supplementary material Text S1. The extraction and the sequencing of the 241 strains collected during the two prospective studies in Denmark was performed as previously described.[19,24]

Bioinformatics methods

All sequencing reads were centralised in Lyon (France) where all bioinformatic analyses were conducted. All these data are available on the European Nucleotide Archive (ENA). The accession numbers can be found in the data availability summary section included in the supplementary material and methods.

Table 1
Distribution of the 924 strains of Staphylococcus aureus in the different cohorts.

Geographical origin	Sequencing country	IE No. of strains (%)	non-IE SAB No. of strains (%)	Total No. of strains	References
AUS/NZ	AUS	49 (*17.88*, 5.30)	438 (*67.38*, 47.40)	487	[22]
DK	DK	120 (*43.80*, 12.99)	121 (*18.62*, 13.10)	241	[15,19]
DK	DK	13 (*4.74*, 1.41)	13 (*2.00*, 1.41)	26	Bacteraemia Repository (SSI)
FR	FR	72 (*26.28*, 7.79)	58 (*8.92*, 6.28)	130	[3]
US	DK	20 (*7.30*, 2.16)	20 (*3.08*, 2.16)	40	[21]
Total		274	650	924	

Percentages in italic or underlined are calculated with respect to the total number of either IE or non-IE SAB strains, or with respect to the total number of strains, respectively.
Abbreviations: IE, infectious endocarditis; SAB, Staphylococcus aureus bacteraemia; AUS, Australia; NZ, New Zealand; DK, Denmark; FR, France; US, United States of America; SSI, Statens Serum Institut.

Quality control

First, raw reads were trimmed in order to remove adapters and other Illumina-specific sequences. Then, a sliding window with an average quality threshold trimming of 20 was processed using Trimmomatic v0.36.[25] Reads whose length was less than 30 nucleotides were removed. Data were subjected to a quality check through FastQC v.0.11.6[26] and summarised using MultiQC v1.3.[27]

Assembly and sequence typing

Assemblies were processed using SPAdes v3.11.1,[28] enabling the automatic coverage cut-off and the mismatch/short indels reducing step. Contigs were broken in k-mers using Kraken v1.1.1[29] in order to remove contigs that did not share at least one k-mer belonging to *S. aureus*. Finally, contigs less than 300 nucleotides in length were manually removed. An assembly quality check was performed using Quast v4.6.1[30] and visualised by R v3.6.3.[31] Sequence types (STs) were determined using MLST v2.9[32,33] software and STs were grouped in CCs.

Single nucleotide polymorphism (SNPs) analysis

Variant calling analysis was performed on the entire data collection using TCH60 (ST30) as the reference genome (GenBank accession nos. **NC_017342.1**). First of all, this genome was aligned against itself using MUMmer v4.0.0[34] in order to locate duplicate regions. Then, Snippy v4.4.5[35] software with a minimum of 10x variant site coverage and a new allele proportion set to 0.5 was used on each dataset, producing a presence/absence matrix for SNP, insertion, or deletion positions for the whole genome. Genomic modifications detected in duplicate regions were removed. The remaining genomic changes were encoded as "1″ when they were different from the reference. By combining the alignment of core SNPs purged of recombinations in nucleotide sequences by Gubbins v2.4.1[36] and FastTree v2.1.9,[37] a phylogenetic tree was inferred with a maximum likelihood method and a generalised time reversible model. This global tree was then visualised using iTOL v5.6.1.[38]

Gene and non-coding RNA analysis

Annotations were performed using Prokka v1.14.5[39] with NCTC8325 (GenBank accession nos. **NC_007795.1**) GenBank annotation file as first protein database. Coding sequences (CDS) identification was reduced to a presence/absence matrix using Roary v3.13.0[40] with the split paralog option enabled. A specific database containing 659 genomic non-coding RNA sequences from five strains of *S. aureus* was constructed as follows: 538 ncRNAs from NCTC8325 (http://srd.genouest.org/) were included in the main database; each ncRNA sequence from other genomes was mapped to the main database using blast v2.2.31[41] with parameters set to 90% identity and coverage; finally, 47, 51, 16, and 7 new genomic sequences were added from N315, JKD6008, Newman, and USA300_FPR3757, respectively. This database was then run through Abricate v0.7[42] with the alignment process set at 90% coverage and identity. The results were transformed into a second presence/absence matrix that was then merged with the presence/absence matrix of the genes.

Genome-wide association study

GWAS was performed using DBGWAS (for *De Bruijn Graph GWAS)* v0.5.2,[43] a method relying on k-mers (all nucleotide substrings of length k found in the genomes) and De Bruijn graphs that connect overlapping k-mers (here DNA fragments), yielding a compact summary of all variations across a set of genomes. Each k-mer presence was calculated both in IE and non-IE SAB samples. In order to take into account the population structure, the phylogenetic tree produced by variant calling analysis and the de novo genome assemblies were used as input to DBGWAS. K-mers were set with 3 different values ([21],[31], and[41]), minor allele frequency filter set to 0.01, and the neighbourhood node number set to 5.

Statistics

All statistical analyses were performed using R v3.6.3.[31] Principal component analysis (PCA) and hierarchical clustering (HC) with complete method and Euclidean distance for both rows and columns were performed to screen for possible stratification of the data. Fisher exact tests were used for proportion comparisons, Wilcoxon-Mann-Whitney tests for distribution comparisons, one sample t-tests to compare the mean of our data to a known value, and Cohen's d to measure the effect size of the difference between two means. A threshold of significant p-value was set at 0.05.

A common heritable trait independent from the genetic background was measured by the Pagel lambda phylogenetic value; in this model, the lambda value varies between 0 (absence of phylogenetic signal on trait distributions) to 1 (strong phylogenetic signal; see Supplementary material Text S1).

The presence or absence of coding sequences and ncRNAs allowing the discrimination of IE strains with non-IE SAB strains was tested using the Least Absolute Shrinkage and Selection Operator (LASSO) regression, accounting for possible bias due to the differ-

ent geographical origins. LASSO regression is a modification of linear regression where the loss of function is modified to minimise the complexity of the model by limiting the sum of the absolute values of the model coefficients (see Supplementary material Text S1). To examine robustness, the Cohen's Kappa coefficient (a quantitative measure of reliability for two raters evaluating the same object, after adjusting this value for what could be expected from chance alone) was calculated from the best model prediction. The value of Cohen's Kappa coefficient was compared to the distribution obtained by the 500 predictions of the random assignment of the observed responses while maintaining the link between the phenotype (IE and non-IE SAB) and the geographical origin. The Cohen's Kappa coefficient can range from −1 (<0, no agreement) to 1 (perfect agreement).

Several R packages were used: ade4 v1.7–18[44], ggplot2 v2.2.1,[45] ComplexHeatmap v2.2.0,[46] factoextra v1.0.7, phytools v0.7–90,[47] permute v0.9–5, glmnet v4.1–2,[48] questionr v0.7.6, effsize v0.8.1,[49] and doParallel v1.0–11.[50]

Rabbit models

Both male (approximately 60%) and female New Zealand white rabbits and weighing 2–3 kg were used. For all experiments, bacterial strains were cultured overnight in Todd-Hewitt broth (Becton Dickinson) to stationary phase and washed in PBS before intravenous inoculation. The combined IE/sepsis model was performed as previously described,[51] i.e. rabbits were injected through the marginal ear veins with either 1–4 × 10^7 or 4.5–5.2 × 10^8 CFUs of *S. aureus* in sterile saline after damage to the aortic valve with a hard-plastic catheter that was then removed. The doses of *S. aureus* used were previously determined to ensure vegetation development within 24 h, increases in vegetation size occurring over the 4-day test period, and to prevent early lethality (death before the 24-h test period). Hence, the following strains were inoculated only at 1–4 × 10^7 CFU/rabbit: ST20101789 and ST20110560 (IE strains), ST20101791 and ST20120211 (non-IE SAB strains), while the four other strains (IE strains: ST20101420, ST20102295; non-IE SAB strains: ST20111372, ST20120206) were inoculated at 4.5–5.2 × 10^8 CFUs. Animals were monitored 4 times per day during experimentations for the development of clinical strokes (defined as the development of limb paralysis requiring premature euthanasia). Experiments performed using rabbits were approved by the University of Iowa Institutional Animal Care and Use Committee (approved animal protocol numbers: 1,106,140 and 4,071,100).

Mice models

To investigate whether IE strains already differed from non-IE SAB strains in initiating the early steps of endocarditis, we examined whether these strains adhered differently on inflamed aortic valves using an early adhesion mouse model as previously described.[52] C57BL/6 wild-type mice were intravenously injected with 2.10^7 CFU fluorescent bacteria. Subsequently, a 32-gauge polyurethane catheter (Thermo Fisher, Waltham, USA; 10 lM) was introduced beyond the aortic valve via the right carotid artery and used to inflame the endothelium via an infusion of histamine (200 mM, infusion rate 10 mL/min for 5 min). Afterwards, mice were immediately sacrificed, hearts were harvested and cryosections of the aortic valve were performed. Adhesion was visualised by confocal microscopy (LSM880, Carl-Zeiss) and quantified using Imaris (Bitplane, Zurich, Switzerland). To analyse the IE development, a long-term inflammation-induced endocarditis model was used as previously described.[52] In this model, mice were either inoculated with 2.10^6 CFU *S. aureus* via the tail vein (producing a low proportion of IE) or inoculated with 4.10^6 CFU *S. aureus* via the transaortic catheter used for histamine infusion (leading to a higher proportion of IE). Mice were monitored for three days to assess the development of endocarditis. The institutional Ethical Committee (KU Leuven) approved all mouse experiments (licence number 189/2017).

Results

Overview of the genomic study

This work investigated genomic elements to distinguish IE from non-IE SAB, focussing on finding i) a common heritable trait associated with IE occurrence from SNP analysis via Pagel's Lambda calculation, ii) a combination of coding sequences (including genes involved in virulence, metabolism, cell-wall synthesis, resistance, etc.…) and/or non-coding RNAs associated with IE occurrence using a LASSO regression, iii) the presence/absence of one or more genomic elements associated with IE occurrence through a Genome-wide association study.

Assemblies and sequence typing

The entire data collection included 274 IE and 650 non-IE SAB samples from four geographical origins. An assembly summary is reported in Fig. S1 and in the Supplementary excel File. A total of 106 STs were identified amongst the 924 *S. aureus* strains in the collection. Each isolate was assigned to one of the 28 CCs (Table S1). The data set was essentially composed of six clonal complexes, namely CC30, CC45, CC8, CC5, CC1, and CC15, each accounting for 15.8 to 8.0% of the entire collection and overall representing 74.2% of this collection. The six largest CCs and the five largest ST distributions amongst the geographical origin are represented in Fig. S2.

SNPs and Pagel lambda phylogenetic analyses

A total of 15,436 SNPs were identified among the 924 *S. aureus* samples using the TCH60 (ST30) as reference (GenBank accession nos. **NC_017342.1**) after duplicated region and recombination corrections. 1127 and 1043 SNPs were present in more than 5% and 10% of the entire data collection, respectively. The phylogeny of all the 924 strains according to infection type, CCs and STs assignment is shown in Fig. 1. Discrimination between IE and non-IE SAB was primary searched through the Pagel lambda phylogenetic signal measurement (Supplementary Text S2) within the entire strain collection and within each of the four geographical origins (Table S2). The lambda (phylogenetic signal) value of the analysis of the global data set reached 0.9441 and was above the one calculated through random permutations (0.7411, 95% CI [0.7198; 0.7624]; *t*-test, p-value < 2.2e-16). Further analysis showed that the unequal ratio of IE/non-IE SAB between the different cohorts biased the model and that ultimately there was no relevant phylogenetic signal explaining the IE and non-IE SAB phenotypes in the whole collection (Supplementary Text S2).

Coding sequence and non-coding RNA analysis

Beyond single nucleotide polymorphisms, the entire coding sequences and non-coding RNAs were tested in search for a discriminant signal between IE and non-IE SAB strains. Amongst the 20,079 different genomic elements that were annotated by Prokka[39] and blasted[41] through the 924 samples, 1002 genes and/or ncRNAs were discarded since they were shared by all samples from the entire strain collection. 13,692 genes were found in less than 5% of the strains and were considered as accessory genes. Finally, 4757 (∼23%) genes and/or ncRNAs were found in more than 5% and less than 95% of the samples and were used for analyses.

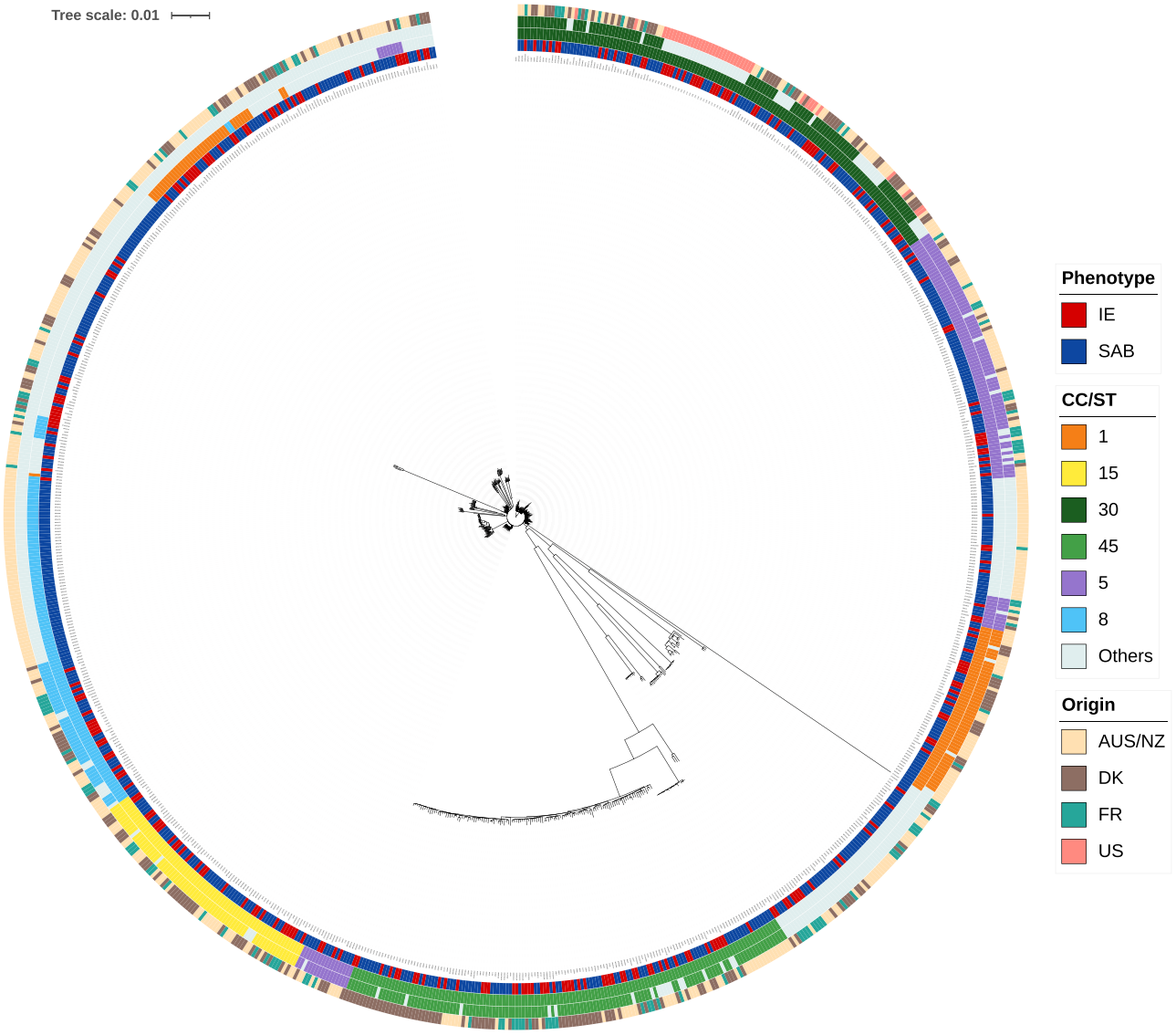

Fig. 1. Rooted Phylogenetic tree of all the 15,436 SNPs from the 924 *S. aureus* strains. The tree is rooted by the TCH60 *S. aureus* strain (ST30; GenBank accession nos. NC_017342.1) using a maximum likelihood method and a generalised time-reversible model after repeated region and recombination corrections. The inner circle represents IE (red) and non-IE SAB (blue) strains. The six largest clonal complexes and five largest sequence types are coloured as follows: CC/ST1, orange; CC/ST15, yellow; CC/ST30, dark green; CC/ST45, light green; CC/ST5, purple; CC/ST8, light blue; other CCs or STs are pooled together (grey). The outer strip refers to the four geographical origins of the samples (AUS/NZ: beige; DK: brown; FR: turquoise; US: pink).

Hierarchical clustering (Fig. 2) and PCA plots (Fig. S4) on the 4757 coding sequences and ncRNAs revealed a major contribution of CC membership to the presence/absence of coding sequences and ncRNAs. Conversely, samples were not distinguished according to their clinical phenotype as it can be seen on the PCA plots (Fig. S4). Of note, similar figures were obtained when restricting the analyses to the 191 virulence genes from the Virulence Finder Database (VFDB) (Fig. S5–6).

The presence or absence of genes and ncRNAs allowing the discrimination of IE strains with non-IE SAB strains was tested by LASSO regression taking into account the various geographical origins (see Methods). Our entire procedure that was computed 1000 times with novel subsampling for training and validation sets reached a Kappa value mean [95% CI] of −0.2242 [−0.2289; −0.2194] (Fig. S7). Although some of these Kappa values (before permutation) are above the Kappa distribution (i.e., Kappa value exceeding the mean of the null distribution) obtained through random permutation (red dots), these values remain far from significance (=1). Therefore, it is not possible to discriminate between IE and non-IE SAB strains based on the presence/absence of genes and non-coding RNAs.

Genome-wide association study

IE and non-IE SAB assemblies were split in k-mers (i.e., sequence substrings of length k) and tagged in order to search specific genomic signatures associated with IE through DBGWAS[43] including a correction for population structure. Nine overlapping k-mer sequences issued from 3 analyses were found statistically significant between the two infection types. These nine-overlapping k-mer sequences were mapped on TCH60 (GenBank accession nos. **NC_017342.1**) and all nine matched between a rRNA-5S and tRNA-Asn (1,092,881 - 1,092,964 bp) genomic region. Each significant k-mer had other matches in TCH60 due to the number of rRNA-5S and tRNA-Asn present in the genome. To further investigate this result, the 9 k-mer overlapping sequences were concatenated to obtain a sequence of 84 nucleotides. This sequence was then blasted against each of the 924 assemblies by setting

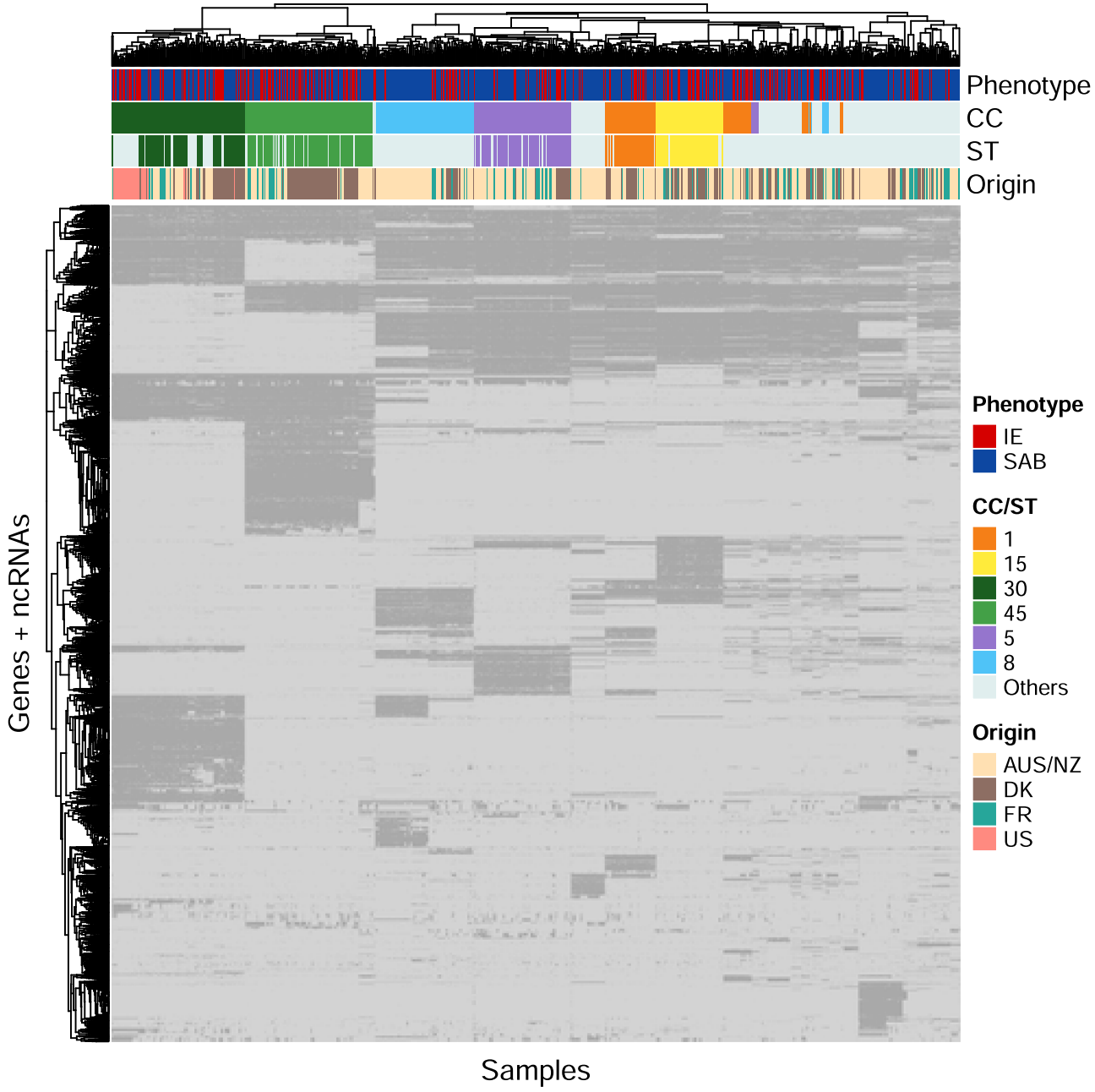

Fig. 2. Presence (dark grey) and absence (light grey) hierarchical clustering of the 4757 coding sequences and non-coding RNAs for the 924 *S. aureus* strains. The blue/red stripe on top shows the infection type, either IE (red) or non-IE SAB (blue). The 2 strips below correspond to the six major clonal complexes and the five major sequence types, respectively (CC/ST1: orange; CC/ST15: yellow; CC/ST30: dark green; CC/ST45: light green; CC/ST5: purple; CC/ST8: light blue; Others: grey). The last strip refers to the four geographical origins of the samples (AUS/NZ: beige; DK: brown; FR: turquoise; US: pink).

the coverage and identity thresholds to 80%. This sequence was extremely prevalent in all strains sequenced at 2 × 250 bp and 2 × 300 bp (345 / 347 strains); it was found at a lower frequency in strains sequenced at 2 × 150 bp (Table S3), which was probably related to differences in sequencing depth and read length between the different cohorts. Indeed, when we tested independently each pair of cohort-reads length against the IE/non-IE phenotypes, none of the associations was significant (Table S4). Overall, this signal could not be considered as significant and must be considered as artefactual. Finally, when the GWAS was repeated on the VIRTSA collection, where almost all strains were sequenced with a read length of 2 × 150 bp, no statistically significant signal associated with IE was observed (data files available at https://sourceforge.net/projects/sab-ie-gwas/files/).

Mice model of infective endocarditis

This model was chosen to test the initial adhesion of *S. aureus* to inflamed aortic valves in mice.[52] Such a model is particularly relevant in the context where we hypothesized that the difference between IE and non-IE SAB strains lies in the early steps of the process when blood-circulating bacteria would seed and develop on cardiac valves. Using a wide screening approach, we chose to pool 13 CC5 IE strains in a single batch and 12 CC5 non-IE SAB strains in another one (Supplementary excel File). CC5 was chosen because it represented the most frequent lineage in the French VIRSTA collection in which IE and non-IE SAB isolates were matched on age and sex in the same geographical setting.[3,17] In a first experiment, we tested whether bacteraemia and endocarditis

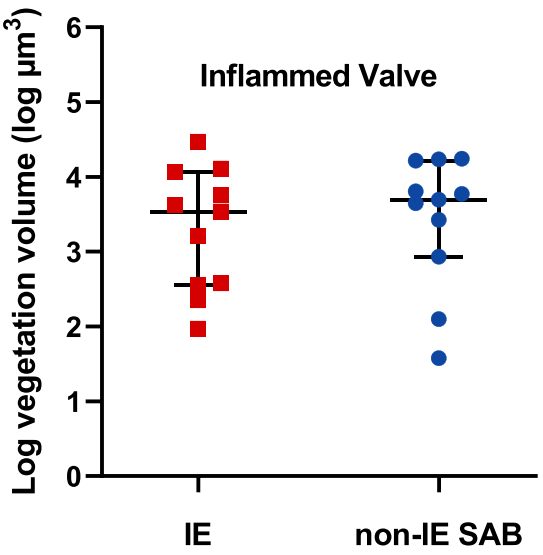

Fig. 3. Adhesion of endocarditis and bacteraemia CC5 strain pools on inflamed cardiac valves. *Staphylococcus aureus* CC5 strain pool (2.10^7 CFU/mouse) were intravenously injected in C57BL/6 mice. Subsequently, via a transaortic catheter, valves were inflamed (5 min of histamine infusion). Mice were immediately sacrificed and adhesion was quantified. Adhesion on inflamed valves of a *S. aureus* CC5 IE strain pool (12 CC5 IE strains, number of mice = 11) was compared with a *S. aureus* CC5 non-IE SAB strain pool (12 CC5 non-IE SAB strains, number of mice = 11, p-value = 0.5619). Results represent log-transformed vegetation volumes in single mice. Median values ± interquartile range are represented; $^*P < 0.05$; Mann-Whitney Wilcoxon test.

strains adhered differently to histamine-inflamed aortic valves. The vegetation volume on the inflamed valves was similar between IE and non-IE SAB strains ($p = 0.5619$; Fig. 3).

We then tested whether the CC5 IE strain pool could cause more IE histologic lesions at Day 3 than the CC5 non-IE SAB strain pool. We observed that both pools equally caused IE on inflamed valves (CC5 OR [95% CI]: 5.0117 [0.8466; 54.4250], p-value = 0.0738; Fig. 4). The challenge was repeated on four CC5 isolates distributed along the phylogeny of the CC5 collection (Supplementary excel File) using the inflammatory model as above and a more severe variation of this model obtained by injecting the inoculum via the transaortic catheter, and there was no significant difference between the IE and non-IE SAB strains (HS OR [95% CI]: 2.1427 [0.5509; 8.8883], p-value = 0.2387, LS OR [95% CI]: Inf. [0.0210; Inf.], p-value = 1; Fig. 5).

The lack of robust evidence with CC5 lineage drove us to test CC45, another common CC in the VIRSTA cohort study. A pool of 10 IE isolates was tested and compared to a pool of 9 non-IE SAB isolates in the inflammatory model. For both pools the proportion of endocarditis was similar (CC45 OR [95% CI]: 2.2475 [0.1055; 144.9300], $p = 0.5996$; Fig. 4).

Rabbit model of infective endocarditis

The above results suggested that the strain potential to initiate IE was similar between IE and non-IE SAB strains. Having established that IE and non-IE strains have a similar ability to initiate IE in the setting of SAB, we sought to investigate whether there were strain differences in the developmental process of IE, such as the size of the vegetation or the blood bacterial load. To this end, we used the mechanical IE-induced rabbit model, a robust experimental model in the field of IE pathogenesis.[51] To optimise the signal/noise ratio, strains were selected within a single CC (CC5, which is one of the most common in the global collection). Four strains of IE and four strains of non-IE SAB were selected for the challenge in rabbits (Supplementary excel File). Upon sacrifice at day 4, all rabbits presented aortic vegetations. The mean vegetation ranged from 15 to 93 mg with no significant difference between strains ($p = 0.0699$, IE vs. non-IE SAB effect size = −0.0277 [−0.6109; 0.5554]) (Fig. 6A). Likewise, the number of viable bacteria in the vegetation was comprised between 1.7×10^7 and 2.6×10^9 CFU (most where in the 10^8 range) with no significant difference between strains ($p = 0.1822$, IE vs. non-IE SAB effect size = 0.2060 [−0.3786; 0.7907]) (Fig. 6B). The number of viable bacteria in the blood ranged from 1.3×10^3 to 2.1×10^5 CFU with no significant difference between them ($p = 0.7356$, IE vs. non-IE

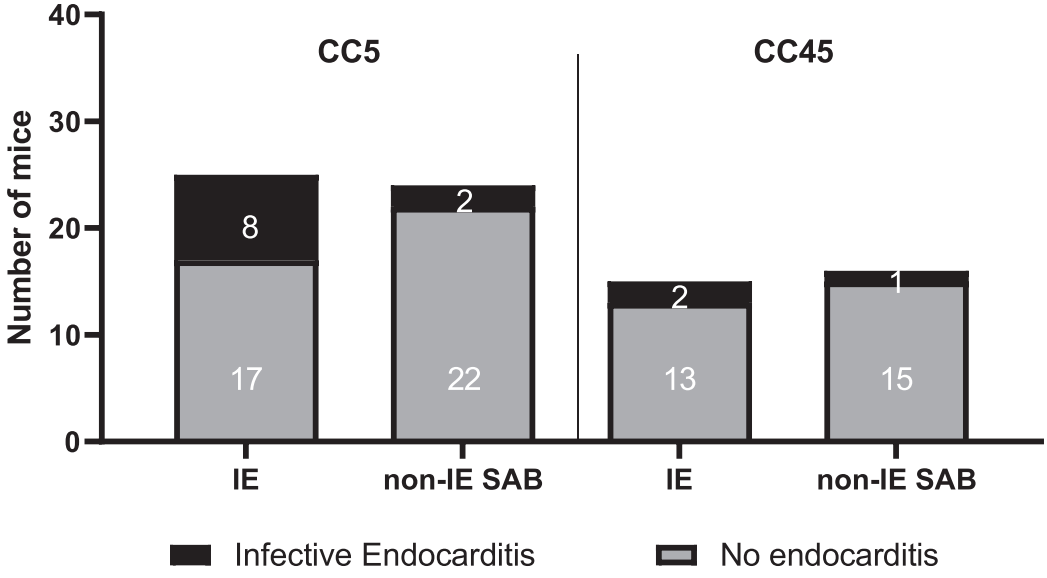

Fig. 4. Propensity of endocarditis and bacteraemia CC5 and CC45 strain pools to cause IE on inflamed cardiac valves. *Staphylococcus aureus* CC5 or CC45 strain pools (2.10^6 CFU/mouse) were intravenously injected in C57BL/6 mice. Subsequently, via a transaortic catheter, valves were inflamed (5 min of histamine infusion). Afterwards, the catheter was removed, mice were monitored for three days to see if endocarditis developed. The proportion of mice that developed endocarditis (i) in the CC5 strain pool vs. no endocarditis (IE strain pool: 12 IE strains, number of mice = 25; non-IE SAB strain pool: 12 non-IE SAB strains, number of mice = 24; OR [95% CI]: 5.0117 [0.8466; 54.4250], p-value = 0.0738); (ii) in the CC45 strain pool vs. no endocarditis (IE strain pool: 10 IE strains, number of mice = 15; non-IE SAB strain pool: 9 non-IE SAB strains, number of mice = 16; OR [95% CI]: 2.2475 [0.1055; 144.9300], p-value = 0.5996).

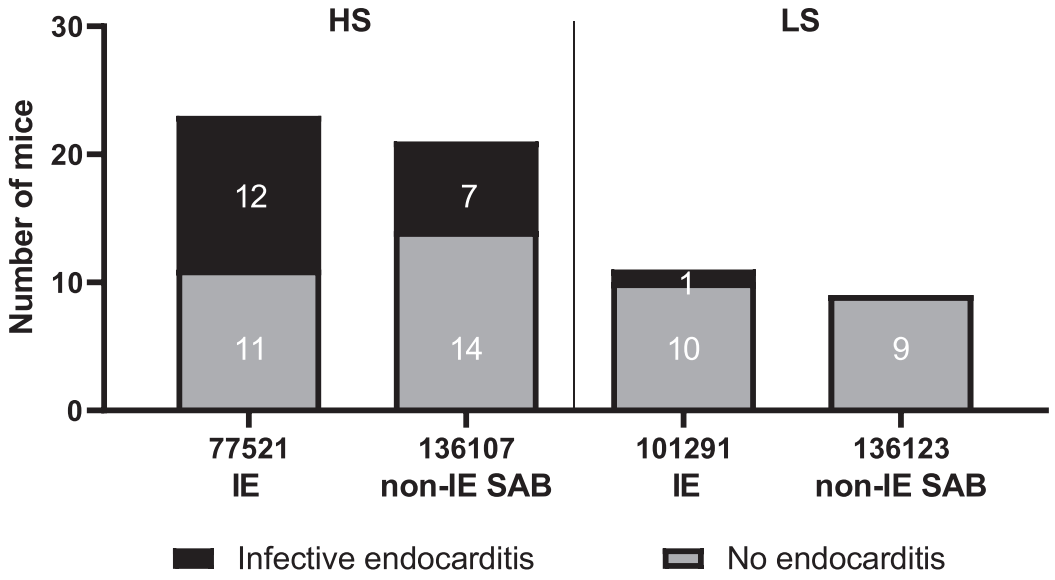

Fig. 5. Propensity of endocarditis and bacteraemia single strains of CC5 to cause IE on inflamed cardiac valves. *S. aureus* IE and non-IE SAB strains (2–4.10^6 CFU/mouse) were perfused via the tail vein catheter (strain 101,291 and 136,123; low severity model, LS) or via a transaortic catheter (increasing the model severity, HS; strain 77,521 and 136,107) in C57BL/6 mice. Valves were inflamed by histamine infusion (5 min) via a transaortic catheter. Afterwards, the catheter was removed and mice were monitored for one (strain 77,521 and 136,107) or three days (strain 101,291 and 136,123) to see whether endocarditis developed. Proportions of endocarditis in mice infected with an IE strain (strain 77,521, number of mice = 23) vs a non-IE SAB strain (strain 136,107, number of mice = 21; HS OR [95% CI]: 2.1427 [0.5509; 8.8883], p-value = 0.2387) and with an IE strain (101,291, number of mice = 11) vs a non-IE SAB strain (strain 136,123, number of mice = 9; LS OR [95% CI]: Inf. [0.0210; Inf.], p-value = 1). Fisher's exact tests ($P^* < 0.05$).

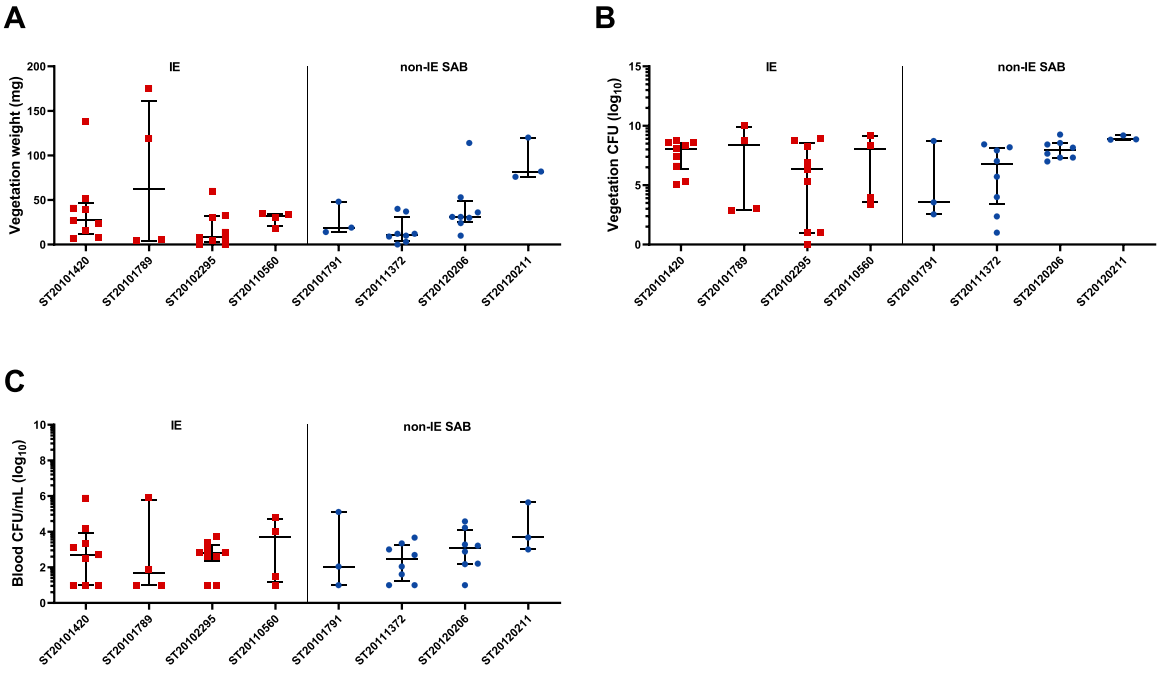

Fig. 6. Native valve, infective endocarditis in rabbits with individual bacteraemia and infective endocarditis strains from the CC5 clonal group. Rabbits were infected intravenously with 1×10^7 - 5×10^8 CFUs/rabbit after mechanical damage to the aortic valves, and infection was allowed to progress for a maximum of 4 days. The doses of *S. aureus* used were previously determined to ensure vegetation development within 24 h, increases in vegetation size occurring over the 4-day test period, and to prevent early lethality. (A) Total weight of vegetations dissected from aortic valves; (B) bacterial counts recovered from aortic valve vegetations shown in panel A; (C) bacterial counts per millilitre of blood at the end of experimentation. Horizontal lines and error bars represent median values with interquartile ranges. There is no statistical difference between strains in terms of vegetation weight, vegetation CFU, and blood CFU. Kruskal-Wallis tests ($P^* < 0.05$; A, p-value = 0.0699; B, p-value = 0.1822; C, p-value = 0.7356).

SAB effect size = 0.2145 [−0.3702; 0.7993]) (Kruskal-Wallis tests) (Fig. 6C) or in blood was higher compared to others.

Altogether, the model of damage-induced infective endocarditis in rabbits showed that all strains, irrespective of their clinical origin in humans (IE or non-IE), induced IE of undistinguishable severity.

Discussion

We herein aimed to address the question of whether specific *S. aureus* strains cause IE in the course of bacteraemia. To ensure a comprehensive answer, we based our approach on (i) various genomic analyses and (ii) two different animal models exploring both

early and late events of aortic valve infections, and both mechanical and inflammation damage-induced IE. All these approaches concordantly revealed that IE isolates could not be discriminated from non-IE SAB isolates based on phylogeny, gene and non-coding RNA content, and k-mers, or in *vivo* virulence.

We analysed 924 *S. aureus* strains of Duke-definite native valve IE and bacteraemia without IE originating from France, Denmark, Australia/New-Zealand, and the US, which constitutes, to our knowledge, the largest collection ever described of such well-defined isolates. The bacteraemia group contained only cases of patients who underwent transthoracic and/or transoesophageal echography, and the endocarditis group included only definite endocarditis on native valves according to the modified Duke criteria. Since we utilised different cohorts from various countries, and enriched for IE cases in some cohorts, there was a risk of geographic bias. However, this potential bias was mitigated either by analysing each cohort independently or by including the geographic origin as a variable. Another result supporting this assertion lies in the genetic diversity of the global collection based on clonal complexes (CC) that is in line with other studies as CC45, CC30, CC5, CC8, CC15, and CC1 were the most prevalent.[11,12,15,18]

Regarding genetic analyses, the bioinformatics issues related to every specific method were overcome by investigating the presence/absence of genes, non-coding RNAs, single nucleotide polymorphisms, and k-mers. Assuming that it is unlikely to find an association between IE with a single genetic event, a combination of markers was screened through the Pagel lambda and LASSO algorithm. We therefore did not confirm the prevalence of specific virulence factor genes in IE strains as previously described by others.[11,12,17] A previous study using microarray data from a small cohort (72 IE and 54 SAB) and discriminant analysis of principal component (DAPC) multivariate analysis method has been able to discriminate IE from non-IE SAB and reassigned IE strains in 80.6% of the cases.[17] The present study performed on a much larger cohort and relying on whole genome sequence data did not confirm these previous observations, which were likely biased by the small sample size of the study.

Although based on over 900 genomes, the number of genetic markers tested was higher than the number of strains, which constitutes a statistical limitation of our study. This limitation was taken into account, in particular in the use of the LASSO algorithm, which eliminates non-essential variables from the final model.

If factors associated with the occurrence of IE do exist, we can conclude that there is no common heritable trait independent from the genetic background; rather, several independent pathways could be at play as an equivalent of “phenotypic convergence”, the phenotype being the propensity to cause IE. Other studies using GWAS to explore the transition from nasal carriage of *S. aureus* to SAB have not identified genomic predictors of bacteraemia versus carriage, perhaps due to the possible polygenic nature of this transition[53]; conversely, another study showed that antimicrobial resistance determinants were associated with bacteremia suggesting that resistance may increase the risk to cause invasive disease.[54] Denamur et al. have estimated that increasing the sample size by one order of magnitude may provide enough power to detect subtle differences in a subset of strains.[55] However, the inclusion of such a high number of patients will be challenging, and if genetic markers are identified, they are likely to be valid only for that one lineage (i.e., a specific ST or CC), and thus difficult to exploit in a diagnostic context. Some concerns in genome-wide association studies are the use of only one reference genome and therefore the risk of missing SNPs on genes not present in this genome. To overcome this bias, we performed k-mers analyses that did not require the use of a reference strain and that accounted for lineage effects in the linear mixed model by including the phylogenetic distance calculated from the variant calling analysis. To reduce the risk of false positives in GWAS, we agree with Young et al. that it would be desirable to use genomes from the same sequencing technology or even the same sequencing run to generate datasets with identical read sizes and relatively similar sequencing depth.

Previous in vitro experiments have reported or hypothesised numerous functions to be involved in staphylococcal IE pathogenesis (resistance to microbicidal peptides, adherence to fibrinogen and fibronectin, biofilm formation, staphylokinase production, platelet aggregation, CD69 superantigen-induced expression, and adhesion to and internalisation by endothelial cells) but have failed to discriminate IE from non-IE SAB strains.[17] To ensure an ultimate “functional” view of the issue, we explored IE animal models, using both murine and rabbit models, as well as both prior mechanical and inflammatory valve damage. Using the murine model developed by Liesenborghs et al.,[52] which allows to study the initial stages of *S. aureus* IE, we showed that all tested *S. aureus* strains could adhere to inflamed valves and induce IE. Using this unique model, no significant difference was observed, using either CC5 or CC45 strains. A weak signal was obtained when testing the CC5 strains in batches but was not confirmed when testing individual strains. We then used the damage-induced IE model in rabbits that showed that all strains, irrespective of their clinical origin in humans (IE or non-IE), induced IE of undistinguishable severity. Moreover, the apparent between-strains variations in vegetation weight, vegetation CFU, and blood CFU never reached significance, suggesting that all strains were equivalent regarding IE severity.

A growing body of evidence suggests that human genetic variability may influence the risk for and severity of *S. aureus* infections, including complicated SAB and IE. Evidence supporting a genetic basis for the susceptibility to SAB include: i) higher rates of *S. aureus* infections in distinct ethnic populations[56-60], ii) familial clusters of *S. aureus* infection[61-64], iii) rare genetic conditions causing susceptibility to *S. aureus*[65-70], and iv) variable susceptibility to *S. aureus* in inbred mice[71-75] and cattle.[76,77] One GWAS has evidenced the human genetic susceptibility to *S. aureus* infection by comparing 4701 cases of *S. aureus* infection and 45,344 matched controls.[78] These results have been reinforced by an admixture mapping study identifying the HLA class II region on chromosome 6 associated with SAB susceptibility.[79] Most recently, investigators have demonstrated a genetic basis for both complicated SAB[80] and persistent MRSA bacteraemia.[81] For example, Mba Medie et al. have employed a carefully matched group of patients with both persistent and resolving MRSA bacteraemia, and found that a single polymorphism in DNMT3A was significantly associated with: i) resolving MRSA bacteraemia, ii) increased global methylation, and iii) reduced levels of IL-10.[81] In 2018, the first published GWAS including 67 patients presenting definite native valve IE (cases) compared to 72 patients with SAB has identified 4 SNPs located on chromosome 3 associated with IE.[82] However, this result did not reach conventional genome-wide significance, possibly due to the small size of the cohorts. This highlights the need to analyse large collections of samples, which can only be obtained within the framework of an international multicentre project.

Another explanation for the lack of association between bacterial or human genetic factors and IE might be that the “IE phenotype” is related to differential (bacterial or human) gene expression or epigenetic changes that cannot be captured through whole-genome sequencing.

One limitation of this study lies in the heterogeneous nature of the various cohorts, notably the ratio of IE to non-IE SAB samples. While the heterogeneity in IE prevalence (and in population structure) is not ideal, it reflects the reality of performing a clinical outcomes GWAS with sufficient statistical power in 2022; this is likely to be easier in the future, for example if clinical metadata are made publicly available along with sequences deposited in public

repositories. As of today, the present study, which compared the largest collection of *S. aureus* bacteraemia and endocarditis strains, indicates that these strains cannot be differentiated on the basis of genomic data or experimental animal models; rather, the conclusion is that all *S. aureus* isolated from blood culture can potentially cause infective endocarditis. In the management of *S. aureus* bacteraemia, for which the indication of ultrasound investigation, in particular TEE, is debated,[83] our results argue for a more systematic investigation of SAB.

Declaration of Competing Interest

FV reports research funding outside the scope of the present study by bioMérieux, two patents pending in antimicrobial resistance detection and shares in Weezion. VGF reports personal fees from Novartis, Novadigm, Durata, Debiopharm, Genentech, Achaogen, Affinium, Medicines Co., Cerexa, Tetraphase, Trius, MedImmune, Bayer, Theravance, Basilea, Affinergy, Janssen, xBiotech, Contrafect, Regeneron, Basilea, Destiny, Amphliphi Biosciences. Integrated Biotherapeutics; C3J, grants from NIH, MedImmune, Cerexa/Forest/Actavis/Allergan, Pfizer, Advanced Liquid Logics, Theravance, Novartis, Cubist/Merck; Medical Biosurfaces; Locus; Affinergy; Contrafect; Karius; Genentech, Regeneron, Basilea, Janssen, from Green Cross, Cubist, Cerexa, Durata, Theravance; Debiopharm, Royalties from UpToDate; and a patent pending in sepsis diagnostics.

Acknowledgments

We thank Hélène Boyer from the *Hospices Civils de Lyon* (France) for help in manuscript preparation.

Funding

SB was supported in part by the French *Agence nationale de la recherche* (ANR IDAREV). WSP was supported by grant R01AI34692-01 from the National Institutes of Health. VGF was supported in part by grant 1R01AI165671 from the National Institutes of Health.

Information on author access to data

The data of this article are available in several different repositories detailed in the Data availability summary supplementary text.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jinf.2022.12.028.

References

1. El Atrouni WI, Knoll BM, Lahr BD, Eckel-Passow JE, Sia IG, Baddour LM. Temporal trends in the incidence of Staphylococcus aureus bacteremia in Olmsted County, Minnesota, 1998 to 2005: a population-based study. *Clin Infect Dis* 2009;**49**(12):e130–8.
2. Rhee Y, Aroutcheva A, Hota B, Weinstein RA, Popovich KJ. Evolving epidemiology of Staphylococcus aureus bacteremia. *Infect Control Hosp Epidemiol* 2015;**36**(12):1417–22.
3. Le Moing V, Alla F, Doco-Lecompte T, Delahaye F, Piroth L, Chirouze C, et al. Staphylococcus aureus bloodstream infection and endocarditis - a prospective cohort study. *PLoS One* 2015;**10**(5):e0127385.
4. Mylonakis E, Calderwood SB. Infective endocarditis in adults. *N Engl J Med* 2001;**345**(18):1318–30.
5. Baddour LM, Wilson WR, Bayer AS, Fowler VG, Tleyjeh IM, Rybak MJ, et al. Infective endocarditis in adults: diagnosis, antimicrobial therapy, and management of complications: a scientific statement for healthcare professionals from the American Heart Association. *Circulation* 2015;**132**(15):1435–86.
6. Habib G, Lancellotti P, Antunes MJ, Bongiorni MG, Casalta JP, Del Zotti F, et al. 2015 ESC guidelines for the management of infective endocarditis: the task force for the management of infective endocarditis of the European Society of Cardiology (ESC) endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). *Eur Heart J* 2015;**36**(44):3075–128.
7. Urja P, Walters RW, Vivekanandan R, Kumar M, Abdulghani S, Hari Belbase R, et al. Trends in the use of echocardiography in patients with Staphylococcus aureus bacteremia: an analysis using the Nationwide Inpatient Sample data. *Echocardiography* 2019;**36**(9):1625–32.
8. Tubiana S, Duval X, Alla F, Selton-Suty C, Tattevin P, Delahaye F, et al. The VIRSTA score, a prediction score to estimate risk of infective endocarditis and determine priority for echocardiography in patients with Staphylococcus aureus bacteremia. *J Infect* 2016;**72**(5):544–53.
9. Palraj BR, Baddour LM, Hess EP, Steckelberg JM, Wilson WR, Lahr BD, et al. Predicting Risk of Endocarditis Using a Clinical Tool (PREDICT): scoring system to guide use of echocardiography in the management of staphylococcus aureus bacteremia. *Clin Infect Dis* 2015;**61**(1):18–28.
10. Kahn F, Resman F, Bergmark S, Filiptsev P, Nilson B, Gilje P, et al. Time to blood culture positivity in Staphylococcus aureus bacteraemia to determine risk of infective endocarditis. *Clin Microbiol Infect* 2021;**27**(9) 1345.e7–1345.e12.
11. Nethercott C, Mabbett AN, Totsika M, Peters P, Ortiz JC, Nimmo GR, et al. Molecular characterization of endocarditis-associated Staphylococcus aureus. *J Clin Microbiol* 2013;**51**(7):2131–8.
12. Nienaber JJC, Sharma Kuinkel BK, Clarke-Pearson M, Lamlertthon S, Park L, Rude TH, et al. Methicillin-susceptible Staphylococcus aureus endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *J Infect Dis* 2011;**204**(5):704–713.
13. Seidl K, Bayer AS, McKinnell JA, Ellison S, Filler SG, Xiong YQ. In vitro endothelial cell damage is positively correlated with enhanced virulence and poor vancomycin responsiveness in experimental endocarditis due to methicillin-resistant Staphylococcus aureus: MRSA-induced endothelial cell damage. *Cell Microbiol* 2011;**13**(10):1530–41.
14. Xiong YQ, Fowler VG Jr, Yeaman MR, Perdreau-Remington F, Kreiswirth BN, Bayer AS. Phenotypic and genotypic characteristics of persistent Methicillin-resistant Staphylococcus aureus bacteremia in vitro and in an experimental endocarditis model. *J Infect Dis* 2009;**199**(2):201–8.
15. Rasmussen G, Monecke S, Ehricht R, Söderquist B. Prevalence of clonal complexes and virulence genes among commensal and invasive Staphylococcus aureus isolates in Sweden. *PLoS One* 2013;**8**(10):e77477.
16. Laabei M, Uhlemann AC, Lowy FD, Austin ED, Yokoyama M, Ouadi K, et al. Evolutionary trade-offs underlie the multi-faceted virulence of Staphylococcus aureus. Schneider DS, editor. *PLoS Biol* 2015;**13**(9):e1002229.
17. Bouchiat C, Moreau K, Devillard S, Rasigade JP, Mosnier A, Geissmann T, et al. Staphylococcus aureus infective endocarditis versus bacteremia strains: subtle genetic differences at stake. *Infect Genet Evol* 2015;**36**:524–30.
18. Tristan A, Rasigade JP, Ruizendaal E, Laurent F, Bes M, Meugnier H, et al. Rise of CC398 lineage of Staphylococcus aureus among infective endocarditis isolates revealed by two consecutive population-based studies in France. Diep BA, editor. *PLoS One* 2012;**7**(12):e51172.
19. Lilje B, Rasmussen RV, Dahl A, Stegger M, Skov RL, Fowler VG, et al. Whole-genome sequencing of bloodstream Staphylococcus aureus isolates does not distinguish bacteraemia from endocarditis. *Microb Genom* 2017;**3**(11). Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5729915/ .
20. Li J.S., Sexton D.J., Mick N., Nettles R., Fowler V.G., Ryan T., et al. Proposed modifications to the duke criteria for the diagnosis of infective endocarditis. 2000;6.
21. Choi SH, Dagher M, Ruffin F, Park LP, Sharma-Kuinkel BK, Souli M, et al. Risk factors for recurrent Staphylococcus aureus bacteremia. *Clin Infect Dis* 2020 ciaa801.
22. Holmes N.E., Turnidge J.D., Munckhof W.J., Robinson J.O., Korman T.M., O'Sullivan V.N., et al. Antibiotic choice may not explain poorer outcomes in patients with Staphylococcus aureus bacteremia and high vancomycin minimum inhibitory concentrations. 2011;8.
23. Holmes NE, Robinson JO, van Hal SJ, Munckhof WJ, Athan E, Korman TM, et al. Morbidity from in-hospital complications is greater than treatment failure in patients with Staphylococcus aureus bacteraemia. *BMC Infect Dis* 2018;**18**(1):107.
24. Rasmussen RV, Høst U, Arpi M, Hassager C, Johansen HK, Korup E, et al. Prevalence of infective endocarditis in patients with Staphylococcus aureus bacteraemia: the value of screening with echocardiography. *Eur J Echocardiogr* 2011;**12**(6):414–20.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**(15):2114–20.
26. Andrews S., Krueger F., Segonds-Pichon A., Biggins L., Krueger C., Wingett S., FastQC. Babraham, UK; 2012.
27. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**(19):3047–8.
28. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**(5):455–77.
29. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;**15**(3):R46.
30. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**(8):1072–5.

31. Development Core Team R. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from *http://www.R-project.org* .
32. Seemann T. MLST [Internet], 2020. Available from: https://github.com/tseemann/mlst
33. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinf* 2010;**11**(1):595.
34. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. Darling AE, editor. *PLoS Comput Biol* 2018;**14**(1):e1005944.
35. Seemann T. Snippy [Internet]. 2020, Available from: https://github.com/tseemann/snippy
36. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;**43**(3) e15-e15.
37. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;**26**(7):1641–50.
38. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;**47**(W1):W256–9.
39. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**(14):2068–9.
40. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;**31**(22):3691–3.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;**215**(3):403–10.
42. Seemann T. ABRicate [Internet], 2020. Available from: https://github.com/tseemann/abricate
43. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, et al. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events. *PLoS Genet* 2018;**14**(11):e1007758.
44. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Soft* 2007;**22**(4). Available from http://www.jstatsoft.org/v22/i04/ .
45. Wickham H., Chang W., Henry L., Pedersen T.L., Takahashi K., Wilke C., et al. ggplot2: create elegant data visualisations using the grammar of graphics [Internet]. 2019 [cited 2019 Sep 9]. Available from: https://CRAN.R-project.org/package=ggplot2
46. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;**32**(18):2847–9.
47. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods Ecol Evol* 2012;**3**(2):217–23.
48. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Soft* 2010;**33**(1). Available from http://www.jstatsoft.org/v33/i01/.
49. Torchiano M. Effsize - a package for efficient effect size computation [Internet]. Zenodo; 2016 [cited 2022 Feb 4]. Available from: https://zenodo.org/record/1480624
50. Ooi H., Corporation M., Weston S., Tenenbaum D. doParallel: foreach parallel adaptor for the 'parallel' package [Internet]. 2019 [cited 2019 Sep 9]. Available from: https://CRAN.R-project.org/package=doParallel
51. Salgado-Pabón W, Breshears L, Spaulding AR, Merriman JA, Stach CS, Horswill AR, et al. Superantigens are critical for Staphylococcus aureus infective endocarditis, sepsis, and acute kidney injury. Johnson EJ, editor. *mBio* 2013;**4**(4) e00494–e00413.
52. Liesenborghs L, Meyers S, Lox M, Criel M, Claes J, Peetermans M, et al. Staphylococcus aureus endocarditis: distinct mechanisms of bacterial adhesion to damaged and inflamed heart valves. *Eur Heart J* 2019;**40**(39):3248–59.
53. Roe C, Stegger M, Lilje B, Johannesen TB, Ng KL, Sieber RN, et al. Genomic analyses of Staphylococcus aureus clonal complex 45 isolates does not distinguish nasal carriage from bacteraemia. *Microb Genom* 2020;**6**(8).
54. Young BC, Wu CH, Charlesworth J, Earle S, Price JR, Gordon NC, et al. Antimicrobial resistance determinants are associated with Staphylococcus aureus bacteraemia and adaptation to the healthcare environment: a bacterial genome-wide association study. *Microbial Genomics* 2021;**7**(11). Available from https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000700 .
55. Denamur E, Condamine B, Esposito-Farèse M, Royer G, Clermont O, Laouenan C, et al. Genome wide association study of Escherichia coli bloodstream infection isolates identifies genetic determinants for the portal of entry but not fatal outcome. Didelot X, editor. *PLoS Genet* 2022;**18**(3):e1010112.
56. Embil J, Ramotar K, Romance L, Alfa M, Conly J, Cronk S, et al. Methicillin-resistant Staphylococcus aureus in Tertiary Care Institutions of the Canadian Prairies 1990-1992. *Infect Control Hosp Epidemiol* 1994;**15**(10):646–51.
57. Hill PC, Birch M, Chambers S, Drinkovic D, Ellis-Pegler RB, Everts R, et al. Prospective study of 424 cases of Staphylococcus aureus bacteraemia: determination of factors affecting incidence and mortality. *Intern Med J* 2001;**31**(2):97–103.
58. Maguire GP, Arthur AD, Boustead PJ, Dwyer B, Currie BJ. Clinical experience and outcomes of community-acquired and nosocomial Methicillin-resistant Staphylococcus aureus in a northern Australian hospital. *J Hosp Infect* 1998;**38**(4):273–81.
59. Kallen AJ. Health care-associated invasive MRSA infections, 2005-2008. *JAMA* 2010;**304**(6):641.
60. Gualandi N, Mu Y, Bamberg WM, Dumyati G, Harrison LH, Lesher L, et al. Racial disparities in invasive Methicillin-resistant Staphylococcus aureus infections, 2005-2014. *Clin Infect Dis* 2018;**67**(8):1175–81.
61. Crum NF, Lee RU, Thornton SA, Stine OC, Wallace MR, Barrozo C, et al. Fifteen-year study of the changing epidemiology of Methicillin-resistant Staphylococcus aureus. *Am J Med* 2006;**119**(11):943–51.
62. Van Der Meer Jos WM, Van Zwet Theda L, Furth R, Weemaes Corry MR. New familial defect in microbicidal function of polymorphonuclear leucocytes. *Lancet North Am Ed* 1975;**306**(7936):630–2.
63. Zimakoff J, Rosdahl VT, Petersen W, Scheibel J. Recurrent Staphylococcal furunculosis in families. *Scand J Infect Dis* 1988;**20**(4):403–5.
64. Oestergaard LB, Christiansen MN, Schmiegelow MD, Skov RL, Andersen PS, Petersen A, et al. Familial clustering of staphylococcus aureus bacteremia in first-degree relatives: a Danish nationwide cohort study. *Ann Intern Med* 2016;**165**(6):390.
65. Faigle W, Raposo G, Tenza D, Pinet V, Vogt AB, Kropshofer H, et al. Deficient peptide loading and MHC class II endosomal sorting in a human genetic immunodeficiency disease: the Chediak-Higashi syndrome. *J Cell Biol* 1998;**141**(5):1121–34.
66. Holland SM, DeLeo FR, Elloumi HZ, Hsu AP, Uzel G, Brodsky N, et al. STAT3 mutations in the hyper-IgE syndrome. *N Engl J Med* 2007;**357**(16):1608–19.
67. Picard C, von Bernuth H, Ghandil P, Chrabieh M, Levy O, Arkwright PD, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine* 2010;**89**(6):403–25.
68. Comeau JL, Lin TJ, Macken MB, Li B, Ku CL, von Bernuth H, et al. Staphylococcal pericarditis, and liver and paratracheal abscesses as presentations in two new cases of interleukin-1 receptor associated kinase 4 deficiency. *Pediatr Infect Dis J* 2008;**27**(2):170–4.
69. Medvedev AE, Lentschat A, Kuhns DB, Blanco JCG, Salkowski C, Zhang S, et al. Distinct mutations in IRAK-4 confer hyporesponsiveness to lipopolysaccharide and interleukin-1 in a patient with recurrent bacterial infections. *J Exp Med* 2003;**198**(4):521–31.
70. Minegishi Y, Saito M, Tsuchiya S, Tsuge I, Takada H, Hara T, et al. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* 2007;**448**(7157):1058–62.
71. Yan Q, Sharma-Kuinkel BK, Deshmukh H, Tsalik EL, Cyr DD, Lucas J, et al. Dusp3 and Psme3 are associated with murine susceptibility to Staphylococcus aureus infection and human sepsis. Skaar EP, editor. *PLoS Pathog* 2014;**10**(6):e1004149.
72. Ahn SH, Deshmukh H, Johnson N, Cowell LG, Rude TH, Scott WK, et al. Two genes on A/J chromosome 18 are associated with susceptibility to Staphylococcus aureus infection by combined microarray and QTL analyses. Wessels MR, editor. *PLoS Pathog* 2010;**6**(9):e1001088.
73. Yan Q, Ahn SH, Medie FM, Sharma-Kuinkel BK, Park LP, Scott WK, et al. Candidate genes on murine chromosome 8 are associated with susceptibility to Staphylococcus aureus infection in mice and are involved with Staphylococcus aureus septicemia in humans. *PLoS One* 2017;**12**(6):e0179033.
74. von Köckritz-Blickwede M, Rohde M, Oehmcke S, Miller LS, Cheung AL, Herwald H, et al. Immunological mechanisms underlying the genetic predisposition to severe Staphylococcus aureus infection in the mouse model. *Am J Pathol* 2008;**173**(6):1657–68.
75. Johnson NV, Ahn SH, Deshmukh H, Levin MK, Nelson CL, Scott WK, et al. Haplotype association mapping identifies a candidate gene region in mice infected with Staphylococcus aureus. *G3 Genes|Genomes|Genetics* 2012;**2**(6):693–700.
76. Griesbeck-Zilch B, Osman M, Ch Kühn, Schwerin M, Bruckmaier RH, Pfaffl MW, et al. Analysis of key molecules of the innate immune system in mammary epithelial cells isolated from marker-assisted and conventionally selected cattle. *J Dairy Sci* 2009;**92**(9):4621–33.
77. Yoshida T, Mukoyama H, Furuta H, Kondo Y, nosuke Takeshima S, Aida Y, et al. Association of BoLA-DRB3 alleles identified by a sequence-based typing method with mastitis pathogens in Japanese Holstein cows. *Anim Sci J* 2009;**80**(5):498–509.
78. DeLorenze G.N., Nelson C.L., Scott W.K., Allen A.S., Ray G.T., Tsai A.L., et al. Polymorphisms in HLA class II genes are associated with susceptibility to Staphylococcus aureus infection in a white population. 2016;8.
79. Cyr DD, Allen AS, Du GJ, Ruffin F, Adams C, Thaden JT, et al. Evaluating genetic susceptibility to Staphylococcus aureus bacteremia in African Americans using admixture mapping. *Genes Immun* 2017;**18**(2):95–9.
80. Scott WK, Medie FM, Ruffin F, Sharma-Kuinkel BK, Cyr DD, Guo S, et al. Human genetic variation in GLS2 is associated with development of complicated Staphylococcus aureus bacteremia. Crawford DC, editor. *PLoS Genet* 2018;**14**(10):e1007667.
81. Mba Medie F, Sharma-Kuinkel BK, Ruffin F, Chan LC, Rossetti M, Chang YL, et al. Genetic variation of DNA methyltransferase-3A contributes to protection against persistent MRSA bacteremia in patients. *Proc Natl Acad Sci USA* 2019;**116**(40):20087–96.
82. Moreau K, Clemenceau A, Le Moing V, Messika-Zeitoun D, Andersen PS, Bruun NE, et al. Human genetic susceptibility to native valve staphylococcus aureus endocarditis in patients with S. aureus bacteremia: genome-wide association study. *Front Microbiol* 2018;**9**. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5893849/.
83. Kouijzer IJE, Fowler VG, Oever JT. Redefining Staphylococcus aureus bacteremia: a structured approach guiding diagnostic and therapeutic management. *J Infect* 2022 S0163445322006417.