RESEARCH   **Open Access**

CrossMark

Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent *Staphylococcus aureus* bacteraemia

Stefano G. Giulieri[1,2,3], Sarah L. Baines[1], Romain Guerillot[1], Torsten Seemann[4,5], Anders Gonçalves da Silva[4], Mark Schultz[4], Ruth C. Massey[6], Natasha E. Holmes[2], Timothy P. Stinear[1] and Benjamin P. Howden[1,2,4*]

Abstract

Background: Large-scale genomic studies of within-host diversity in *Staphylococcus aureus* bacteraemia (SAB) are needed to understanding bacterial adaptation underlying persistence and thus refining the role of genomics in management of SAB. However, available comparative genomic studies of sequential SAB isolates have tended to focus on selected cases of unusually prolonged bacteraemia, where secondary antimicrobial resistance has developed.

Methods: To understand bacterial genetic diversity during SAB more broadly, we applied whole genome sequencing to a large collection of sequential isolates obtained from patients with persistent or relapsing bacteraemia. After excluding genetically unrelated isolates, we performed an in-depth genomic analysis of point mutations and chromosome structural variants arising within individual SAB episodes.

Results: We show that, while adaptation pathways are heterogenous and episode-specific, isolates from persistent bacteraemia have a distinctive molecular signature, characterised by a low mutation frequency and high proportion of non-silent mutations. Analysis of structural genomic variants revealed that these often overlooked genetic events are commonly acquired during SAB. We discovered that IS*256* insertion may represent the most effective driver of within-host microevolution in selected lineages, with up to three new insertion events per isolate even in the absence of other mutations. Genetic mechanisms resulting in significant phenotypic changes, such as increases in vancomycin resistance, development of small colony phenotypes, and decreases in cytotoxicity, included mutations in key genes (*rpoB*, *stp*, *agrA*) and an IS*256* insertion upstream of the *walKR* operon.

Conclusions: This study provides for the first time a large-scale analysis of within-host genomic changes during invasive *S. aureus* infection and describes specific patterns of adaptation that will be informative for both understanding *S. aureus* pathoadaptation and utilising genomics for management of complicated *S. aureus* infections.

Keywords: *Staphylococcus aureus*, Bacteraemia, Genomics, Within-host diversity, Persistence

* Correspondence: bhowden@unimelb.edu.au
[1]Department of Microbiology and Immunology, The University of Melbourne at the Doherty Institute for Infection & Immunity, Melbourne, Australia
[2]Infectious Disease Department, Austin Health, Melbourne, Australia
Full list of author information is available at the end of the article

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 2 of 17

Background

The outcome of *Staphylococcus aureus* bacteraemia (SAB) is a result of a complex interaction of host, pathogen, and treatment factors. Persistence, usually defined as bacteraemia of greater than 3–7 days duration, is an important factor in SAB outcome [1], including secondary antibiotic resistance development, metastatic infectious complications, and mortality [2]. Persistent bacteraemia involves a sequence of events, including invasion, immune evasion, and establishment of secondary infectious foci, usually all in the context of antimicrobial treatment [3]. From the bacterial perspective, invasive *S. aureus* isolates are subjected to the pressures of the immune response, lack of nutrients, and antibiotics. These environmental challenges constitute a significant selective pressure driving adaptive evolution in the pathogen, and access to sequential isolates from patients with persistent SAB offers the opportunity to understand pathoadaptation during invasive *S. aureus* infections.

Over the last decade, with increasing availability of whole-genome sequencing, within-host genomic studies have addressed *S. aureus* niches that are important for pathogenesis [4]. Studies of colonising isolates have uncovered the "cloud of diversity" of *S. aureus* colonising the host and improved our ability to track transmission networks [5]. Other authors have revealed mutations associated with the evolution from colonising to invasive strain in a single patient [6] or in large cohorts [7]. Genomic studies of sequential blood isolates in persistent SAB have primarily focused on cases with significant phenotypic changes that might arise in persistent infection, such as secondary resistance to antibiotics (especially the vancomycin intermediate phenotype [8, 9], daptomycin resistance [10]), development of the small-colony phenotype [11, 12], or genetic changes associated with extreme cases of persistence [13]. However, these analyses were restricted to a small number of selected cases and thus offer only limited insights on the general pattern of *S. aureus* evolution during SAB. Understanding the typical pattern of within-host evolution during SAB through large-scale investigation of paired isolates will potentially identify shared genomic signatures associated with *S. aureus* adaptation in vivo and inform the use of whole-genome sequencing in the management of SAB more broadly. For example, genomic monitoring of SAB could be used to distinguish true relapses from reinfection with a closely related strain, or track mutations associated with persistence or resistance early in the course of the disease, an approach that has been recently demonstrated for lung cancer [14].
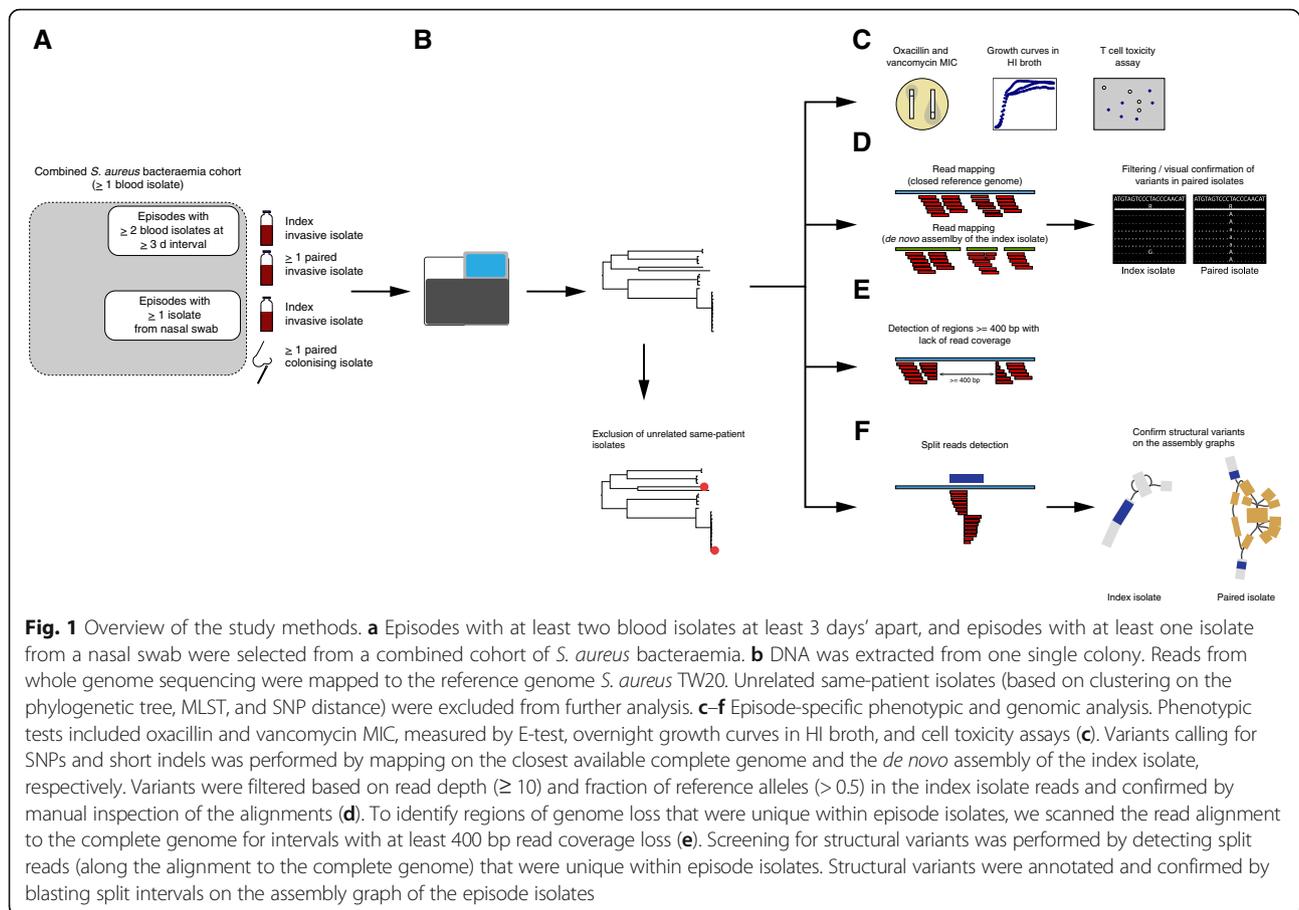
To explore genetic changes associated with persistent or relapsing SAB and compare them to those occurring between colonising and invasive isolates, we applied bacterial whole-genome sequencing to a large cohort of SAB, regardless of phenotypic changes. In addition to the commonly investigated mutational variants, we performed a detailed analysis of chromosome structural variants (e.g. large deletions and insertions, insertions of mobile genetic elements) within same-patient strains. This explorative approach uncovered a diverse mutational landscape and a molecular signature distinctive of persistent bacteraemia. Furthermore, we demonstrate for the first time that structural variation represents an important mechanism promoting genetic plasticity within the host, even in the absence of point mutations and insertions and deletions.

Methods

Case and isolate selection

Isolates included in this study were selected from two multicenter cohorts of SAB (Fig. 1a). The vancomycin substudy of the Australian and New Zealand Cooperative on Outcome in Staphylococcal Sepsis (ANZCOSS) study was a retrospective study of *S. aureus* isolates collected between 2007 and 2008 [15–19]. The Vancomycin Efficacy in Staphylococcal Sepsis in Australasia (VANESSA) cohort was a prospective, multicentre study that has been designed to establish the impact of host, pathogen, and antimicrobial factors on outcome from SAB and has recruited patients between 2012 and 2013 [20]. Both studies collected data on patient demographics, comorbidities, clinical characteristics, duration of bacteraemia, 30-day mortality, and SAB recurrence. Antistaphylococcal treatment of the index SAB episode was also recorded and is listed in Additional file 1: Table S1. Overall, methicillin-resistant *S. aureus* (MRSA) bacteraemia was treated with vancomycin, and methicillin-susceptible *S. aureus* (MSSA) bacteraemia was treated with flucloxacillin or cefazolin (vancomycin was rarely used if there was a contraindication to beta-lactams). Median duration of treatment was 26 days (IQR 14–48). Isolates from subsequent positive blood cultures and from nasal colonisation screening were available for a subset of SAB episodes. Therefore, SAB episodes with at least two blood isolates collected at a minimum of 3 days apart were included as *invasive* episodes, and the isolate collected at the detection of bacteraemia was defined as *index* isolate; blood isolates collected subsequently were defined as *paired invasive*. Episodes for which colonising *S. aureus* isolates were available were included as *colonising-invasive pairs* (*index* isolate: first detected blood isolate; *paired colonising* isolate). Invasive isolates collected after the index isolate were classified according to the clinical context in (i) *persistent bacteraemia* (no negative blood cultures before the collection of the paired isolate); (ii) *relapse on treatment* (at least one negative blood

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 3 of 17

Fig. 1 Overview of the study methods. **a** Episodes with at least two blood isolates at least 3 days' apart, and episodes with at least one isolate from a nasal swab were selected from a combined cohort of *S. aureus* bacteraemia. **b** DNA was extracted from one single colony. Reads from whole genome sequencing were mapped to the reference genome *S. aureus* TW20. Unrelated same-patient isolates (based on clustering on the phylogenetic tree, MLST, and SNP distance) were excluded from further analysis. **c–f** Episode-specific phenotypic and genomic analysis. Phenotypic tests included oxacillin and vancomycin MIC, measured by E-test, overnight growth curves in HI broth, and cell toxicity assays (**c**). Variants calling for SNPs and short indels was performed by mapping on the closest available complete genome and the *de novo* assembly of the index isolate, respectively. Variants were filtered based on read depth (≥ 10) and fraction of reference alleles (> 0.5) in the index isolate reads and confirmed by manual inspection of the alignments (**d**). To identify regions of genome loss that were unique within episode isolates, we scanned the read alignment to the complete genome for intervals with at least 400 bp read coverage loss (**e**). Screening for structural variants was performed by detecting split reads (along the alignment to the complete genome) that were unique within episode isolates. Structural variants were annotated and confirmed by blasting split intervals on the assembly graph of the episode isolates

culture between the index isolate and the paired isolate and collection before the end of antistaphylococcal treatment of the index episode); (iii) *relapse after treatment* (collection after the end of antistaphylococcal treatment of the index episode).

The first blood culture isolate from each episode (index isolate), isolates from blood cultures collected at least 3 days after the index (paired isolates) and colonising isolates were stored at – 80 °C. Phenotypic confirmation of *S. aureus* was performed using the coagulase and DNase tests.

Whole genome sequencing

Bacterial isolates stored in glycerol broth at – 80 °C were subcultured twice onto horse blood agar. Genomic DNA was extracted from single colonies using the Janus® automated workstation (PerkinElmer) or manually using Invitrogen PureLink genomic DNA kit or the Sigma GenElute kit. DNA concentration was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies) and normalised to a concentration of 0.2 ng/μl for library preparation with Nextera® XT DNA (Illumina). Genome sequencing was carried out on the MiSeq® and NextSeq® (Illumina) platforms with a read length of 2 × 150 bp or

2 × 250 bp (Fig. 1b). The quality of sequencing was evaluated by calculating mean read depth (based on a genome length of 3 million bp), and assessing assembly metrics obtained using SPAdes, version 3.9.0 [21]. Quality metrics are listed in Additional file 1: Table S1. Species was confirmed by k-mer classification using Kraken, version 0.10.5-beta [22].

Multi-locus sequence typing and resistome

De novo assemblies of the isolates were generated with SPAdes [21]. Assembled genomes were scanned for MLST typing using MLST, version 2.7 (T. Seemann, https://github.com/tseemann/mlst). Resistance genes were detected from assemblies using Abricate, version 0.3 (T. Seemann, https://github.com/tseemann/abricate) using the ResFinder database [23]. Clonal complexes were inferred using eBurst, version 3 [24].

Global core genome alignment

To obtain a global alignment of all isolates included in the study (both invasive and colonising), sequence reads were mapped to *S. aureus* TW20, a clonal complex (CC) 8/sequence type (ST) 239 methicillin-resistant *S. aureus* (MRSA) reference genome (Fig. 1b). [25]. Read mapping,

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 4 of 17

variant calling, and core genome alignment were performed using the Snippy pipeline, version 3.0 (T. See-mann, https://github.com/tseemann/snippy). Maximum likelihood phylogeny was obtained using IQ-TREE, version 1.6 [26, 27]. Branch support was calculated using both ultrafast bootstrap support [28] and the SH-like approximate likelihood ratio test [29] with threshold values of 95% and 80%, respectively. The phylogenetic tree was rooted using *Staphylococcus argenteus* as an outgroup [30], and plotted and annotated with the R packages *ape* [31] and *ggtree* [32].

Determination of genetic relatedness of same-patient isolates

To establish relatedness between same-patient isolates, we considered tree topology, MLST, and the pairwise single-nucleotide polymorphism (SNP) distance matrix, computed from the core genome alignment using the common reference *S. aureus* TW20 and from the alignment constructed using the *de novo* assembly of index isolate. Only related same-patient isolates were kept for further phenotypic and genomic analysis (episode-specific analysis).

Episode-specific analysis

Figure 1c–f illustrates phenotypic and genomic analyses performed on genetically related isolates collected from the same patient.

Phenotypic testing

Vancomycin and oxacillin minimum inhibitory concentration (MIC) were assessed using Etest (bioMerieux), according to manufacturer's instructions. For growth curves, isolates freshly subcultured were grown overnight in heart infusion (HI) broth, inoculated into 200 μl of fresh HI at a 1:400 dilution, and incubated at 37 °C with agitation during 16 h. Optical density at 600 nm was measured at 15-min intervals using the EnSight™ Multimode Plate Reader (PerkinElmer).

Cytotoxicity assays were performed on a subset of isolates that were selected using the following criteria: bacteraemia duration of at least 7 days, relapse on anti-staphylococcal treatment, vancomycin MIC increase or development of small-colony phenotype, and possible change in toxicity based on genetic changes (e.g. *agr* mutations).

Cytotoxicity was measured using a modified method of that described previously [33, 34]. A single bacterial colony was inoculated into 5 mL brain heart infusion (BHI) broth (Oxoid) and incubated for 18 h at 37 °C with agitation (180 rpm). A 1 mL aliquote was centrifuged (10 min, 10,000 rpm) and the supernatant collected and frozen at − 20 °C. Instead of T2 cells, a THP-1 human monocyte cell line was used. THP-1 cells were cultured in RPMI (Lonza) supplemented with 10%

fetal calf serum, and 1% L-Glutamine (200 mM)−Penicillin (10,000 units)−Streptomycin (10 mg/mL) solution (Sigma) at 37 °C for 2 to 4 days. For testing, THP-1 cells were centrifuged (10 min, 1200 rpm, 22 °C) and resuspended in Dulbecco's Phosphate Buffered Saline (Thermo Scientific) to a concentration of 2.4–3.0 million cells/ml. Bacterial supernatant (diluted 50% in BHI broth) and THP-1 cells were mixed in a 1:1 ratio (using 20 μl volumes) and incubated for 12 min at 37 °C. After incubation, 20 μl of Trypan Blue Solution (Corning) was added and 20 μl was loaded onto a disposable cell counting slide (Immune Systems). Three 4 × 4 grids were counted and averaged to gain a viable- and total-cell count. Each bacterial supernatant was tested in technical duplicate. In the case of the paired invasive isolates (BPH3706 and BPH3757) where a difference in cytotoxicity was detected, two additional biological replicates (each tested in technical duplicate) were performed.

Variant calling

We used Snippy to map sequence reads from the same episode to the closest available complete genome in the NCBI repository (strain names and accession numbers are listed in Additional file 1: Table S1) and to the *de novo* assembly of the index isolate, which was generated using SPAdes as described above and annotated with Prokka [35]. A consensus sequence of the references was generated by mapping the reads of the index isolate (using Snippy). In addition, we filtered variants called for paired isolates by reviewing the alignment of the index isolate reads and excluding positions, where read coverage was below 10 and the fraction of reads identical to the reference was below 50%. All variants were manually validated by comparing the read alignments of the index isolate and the paired isolate. The filtering was performed using SAMtools mpileup, version 1.4, and the alignments were inspected using SAMtools tview, version 1.4 [36].

To test whether masking repetitive regions might improve variant calling, we used NUCmer, version 3.1 [37], to identify repetitive regions on the closest available complete genome and repeated the Snippy run after masking these regions from the reference.

Episode-specific phylogenies

For two cases where multiple isolates per patient were available, we inferred an episode-specific phylogeny. Core-genome alignment on the closest complete genome was "curated" by filtering SNPs as described above. Maximum-likelihood phylogeny was inferred using IQ-tree and tree were rooted at midpoint.

Annotation and functional classification of mutated gene products

The clustering tool CD-HIT, version 4.6.7 [38], was applied to compare proteins whose sequence was altered by mutations confirmed by the approach described above. Unique protein sequences were annotated by assigning Clusters of Orthologous Groups (COGs) using Reverse Position Specific BLAST (rpsblast), version 2.5.0. The COG database was downloaded from NCBI (ftp://ftp.ncbi.nih. gov/pub/mmdb/cdd/little_endian). Rpsblast results were parsed using the Biopython package [39] Blast, module NCBIXML.

Annotation and convergence analysis of intergenic regions

Mutations in intergenic regions were annotated using an adapted Biopython script (http://biopython.org/wiki/Intergenic_regions) and convergence among mutated intergenic regions was explored first using CD-HIT and then by performing an all-vs-all blastn search, as described in the pipeline Piggy [40].

Detection of chromosome structural variants

To detect larger deletions, the episode-specific alignment to the complete reference genome was analysed using BEDTools, version 2.26.0 [41] to identify unique intervals within the patient isolates (i.e. not present in all isolates from the same patient) with at least 400 bp read coverage loss. The 400-bp threshold was determined empirically based on the distribution of the length of deletions using a 1000-bp window in BEDTools coverage. Insertions and smaller deletions were identified by analysis of split reads. Split reads (i.e. reads that cannot be represented by a linear alignment and therefore have one or more supplementary alignments as specified in the SAM format available at https://github.com/samtools/hts-specs) were extracted from the episode-specific alignment to the complete reference genome using a python script that is part of the LUMPY framework (https://github.com/arq5x/lumpy-sv). We kept split reads that were unique to one or more isolate per episode and had a breakpoint (defined by start or stop of the read alignment) with a coverage > 10. Breakpoints and read coverage at breakpoint were obtained by parsing the SAMtools mpileup output. Unique split read intervals were confirmed by manual inspection of the alignment, and the primary interval and the supplementary intervals were annotated using the complete genome in GFF format and BEDTools. To confirm variants identified with the split read analysis, we performed a BLAST search for primary and supplementary intervals on the *de novo* assembly graph of the isolates using Bandage, version 0.8.1 [42]. Structural variants were visualised using Geneious, version 8.1.7 (Biomatters).

Pan-genome analysis

We used Roary, version 3.12.0 [43], to generate ortholog clustering from assemblies of closely related isolates (annotated with Prokka). To exclude genes inaccurately attributed to the accessory genome, we mapped reads to a multi-fasta file of the accessory genes using BWA-MEM, version 0.7.17-r1188 [44], and scanned the alignment with BEDTools. Only genes for which at least one isolate had a coverage breadth of 0 and at least one isolate had a mean coverage depth of 10 were classified as true accessories. The relevance of the resulting accessory genes was checked using blast to confirm that the sequences were *S. aureus*.

IS256 BLAST search

To obtain the clonal distribution of IS*256*, we performed a blastn search using the IS*256* fasta sequence as a query (downloaded from ISfinder https://www-is.biotoul.fr/) and with the following parameters: minimum coverage 90%, minimum identity 95%, wordsize 32, and evalue 0.01. We searched 124 complete genomes available in NCBI repository in May 2017 and 130 draft assemblies of the isolates included in this study.

Statistical analyses

Statistical analyses were performed in R, version 3.4.1. The chi-square test was used to compare proportions of isolates with at least one mutation and proportion of non-silent mutation among isolate groups. Differences in number of mutations among isolate groups were assessed using the Kruskal-Wallis test. Doubling time and maximum grow rate were calculated by fitting curves using local polynomial regression fitting as performed by the R package *cellGrowth* [45]. Enrichment analysis of functional categories among mutated gene products was performed in R by computing the hypergeometric test for each category using the reference genome *S. aureus* TW20 as control.

Results

Population structure of paired isolates from *S. aureus* bacteraemia reveals a broad genetic background

A collection of 130 *S. aureus* isolates from 57 patients was assembled from two multicenter cohorts of SAB and underwent whole genome sequencing, phylogenetic analyses, phenotypic comparisons, and analysis for genomic variants (Fig. 1a–f and Additional file 2: Figure S1). We included 50 SAB episodes with at least two blood isolates collected at a minimum of 3 days apart (50 index isolates, 61 paired invasive isolates). In addition, 12 colonising isolates were collected from 4 episodes already included in the invasive group and 7 supplementary episodes. The median sample delay between index isolate and paired invasive isolate was 8 days (interquartile

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 6 of 17

range [IQR] 5–23). The clinical context of the paired invasive isolate was persistent bacteraemia ($n = 31$), relapse on treatment ($n = 10$), and relapse after treatment ($n = 10$). Among colonising strains, seven were collected before or at the same time as the index sample (median sampling delay 13 days before index, IQR 1.5–71.5) and five were collected afterwards (median delay 5 days, IQR 1–8).

The maximum-likelihood phylogeny of the collection was inferred from 103,974 core genome SNPs (size of the core genome alignment 1,977,743 bp) and is shown in Fig. 2a. The population was dominated by CC8, which represented 36% of isolates, followed by CC45 (14%), CC5 (13%), CC22 (8%), and CC39 (7%). The dominant clade was ST239, including four closely related strains with novel ST types that are single-locus variants of ST239. This clade accounted for 28% of the isolates. This diverse population of *S. aureus* shows that the paired isolates were selected from a broad genetic background.

We then calculated the pairwise SNP distance and used phylogenetic clustering to infer relatedness of paired isolates and thus distinguish between persistent or relapsing bacteraemia and co-infection or reinfection with an unrelated strain (Fig. 1b). We also investigated the genetic distance between index blood isolates and their paired colonising isolate to identify SAB episodes that were unrelated to the sampled colonising isolate. Most same-patient isolates clustered together and exhibited a pairwise SNP distance below 100 (Fig. 2b, c). In this group, pairwise distances ranged between 0 and 98 and 0–25 SNPs, when using the common reference *S. aureus* TW20 and the *de novo* assembly of the index isolate, respectively. Isolates from patient 37 were initially separated by 717 SNPs when mapping to *S. aureus* TW20. However, they were considered genetically related since they clustered together on the tree, both belonged to ST93, and pairwise SNP distance when mapping to the *de novo* assembly of the index isolate was 6. Pairwise SNP distance between patient 19 isolates was 266; however, this was due to phage recombination (see below) and based on tree topology and MLST these isolates were also categorised as genetically related. Nine paired isolates (seven paired invasive and two paired colonising) had a SNP distance to the index larger than 10,000 bp and were also different by multilocus sequence type (MLST). We therefore defined these isolates as genetically unrelated to the index and excluded them from further analysis of in vivo diversity. The seven unrelated paired invasive isolates were collected after a longer interval as compared to isolates that were genetically close to the index sample (median sampling delay 72 vs. 7 days, $p = 0.002$, Fig. 3a). Thus, reinfection with a different clone as defined by genetic unrelatedness occurred in 7 out of 50 (14%) cases of SAB included in this study. This is consistent with a previous publication by Fowler et al., where 20% of SAB recurrences were reinfections, as defined by pulsed-field gel electrophoresis (PFGE), a technique that has lower resolution than whole-genome sequencing (WGS) [46].

After exclusion of unrelated isolates pairs, 51 episodes with 115 isolates were retained for in-depth genomic and phenotypic within-host diversity analysis (Fig. 1c–f).

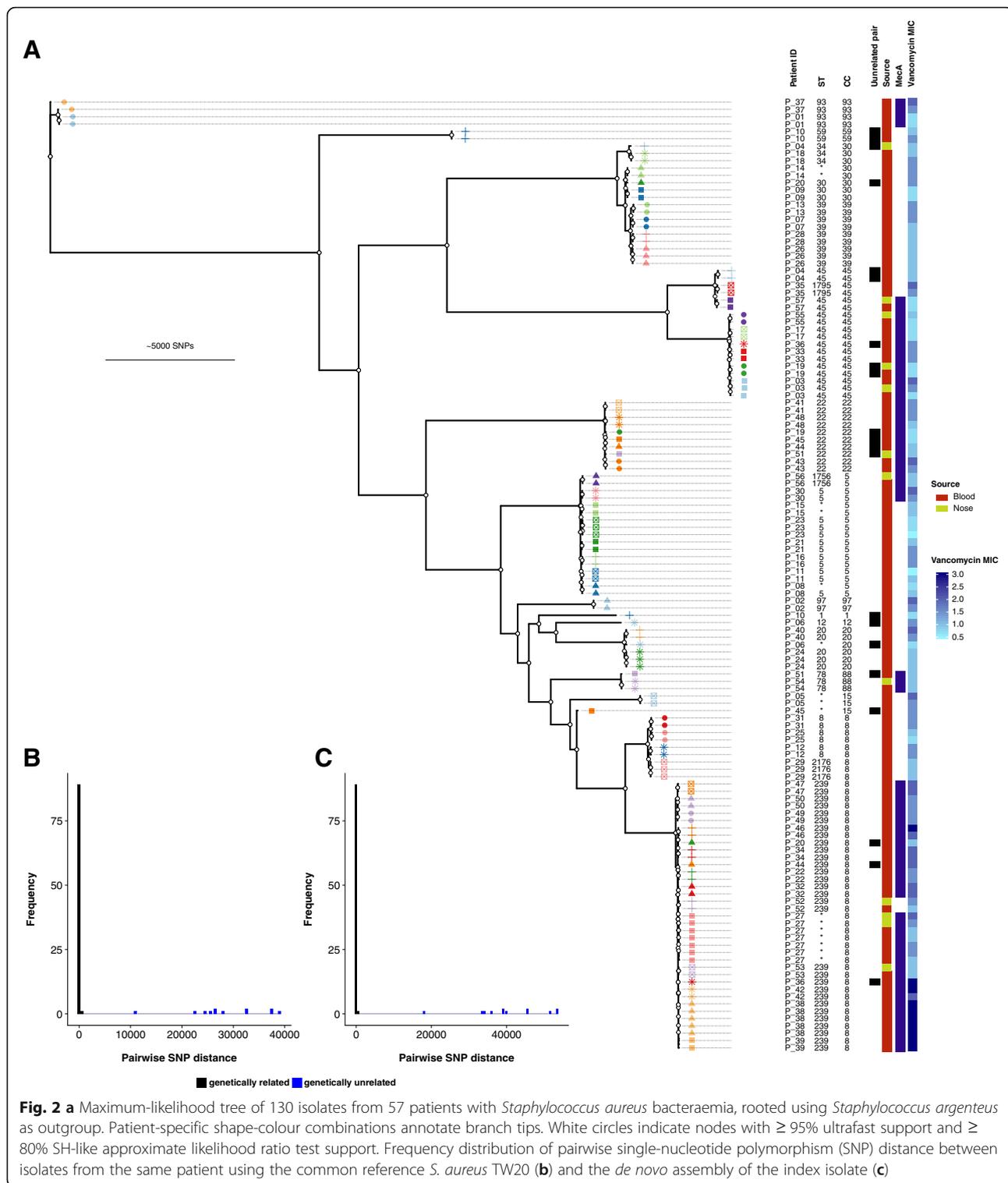## Paired invasive isolates have low genetic diversity

The choice of the reference genome and the filtering of variants has an impact on the number of identified mutations [47]. Therefore, to obtain the most accurate estimate of within-host diversity in paired isolates, we applied an episode-specific genome mapping approach. By mapping sequence reads to both the closest available complete genome from the NCBI repository and a *de novo* polished assembly of the index isolate and thorough review of the variants through manual inspection of the alignments, we were able to effectively eliminate a significant number of false-positive mutation calls and retain only true genetic variation (Additional file 2: Figures S2 and S3). While manual inspection has been applied in other within-host genomic studies [5], one alternative approach may be masking repetitive regions, where mapping errors are more likely to occur, as implemented in some variant calling pipelines [48]. However, using manual inspection as the gold standard, we calculated in our dataset that masking repetitive regions on the reference (closest available complete genome) would result in a sensitivity and specificity of 93% and 21%, respectively.

Using this approach, we identified a total of 182 variants (141 SNPs and 41 small indels [median size 1 bp, IQR 1–1]) in 32 out of 64 paired isolates. We observed very limited genetic diversity in paired invasive isolates compared to paired colonising isolates (Fig. 3b). Only 23 (43%) of 54 paired invasive isolates exhibited at least one mutation, while 9 out of 10 paired colonising isolates were mutated ($p = 0.016$). Among isolates with at least one mutation, the median number of variants in paired invasive and paired colonising isolates was 2 and 12, respectively ($p = 0.014$).

Among 158 unique variants, 81 (51%) were predicted to result in changes in protein function: 60 were missense substitutions, 5 were nonsense substitutions (leading to a premature stop codon), and 16 were frameshift mutations. The remaining 77 mutations occurred in non-coding regions (47) or were synonymous substitutions (30) (Additional file 3: Table S2).
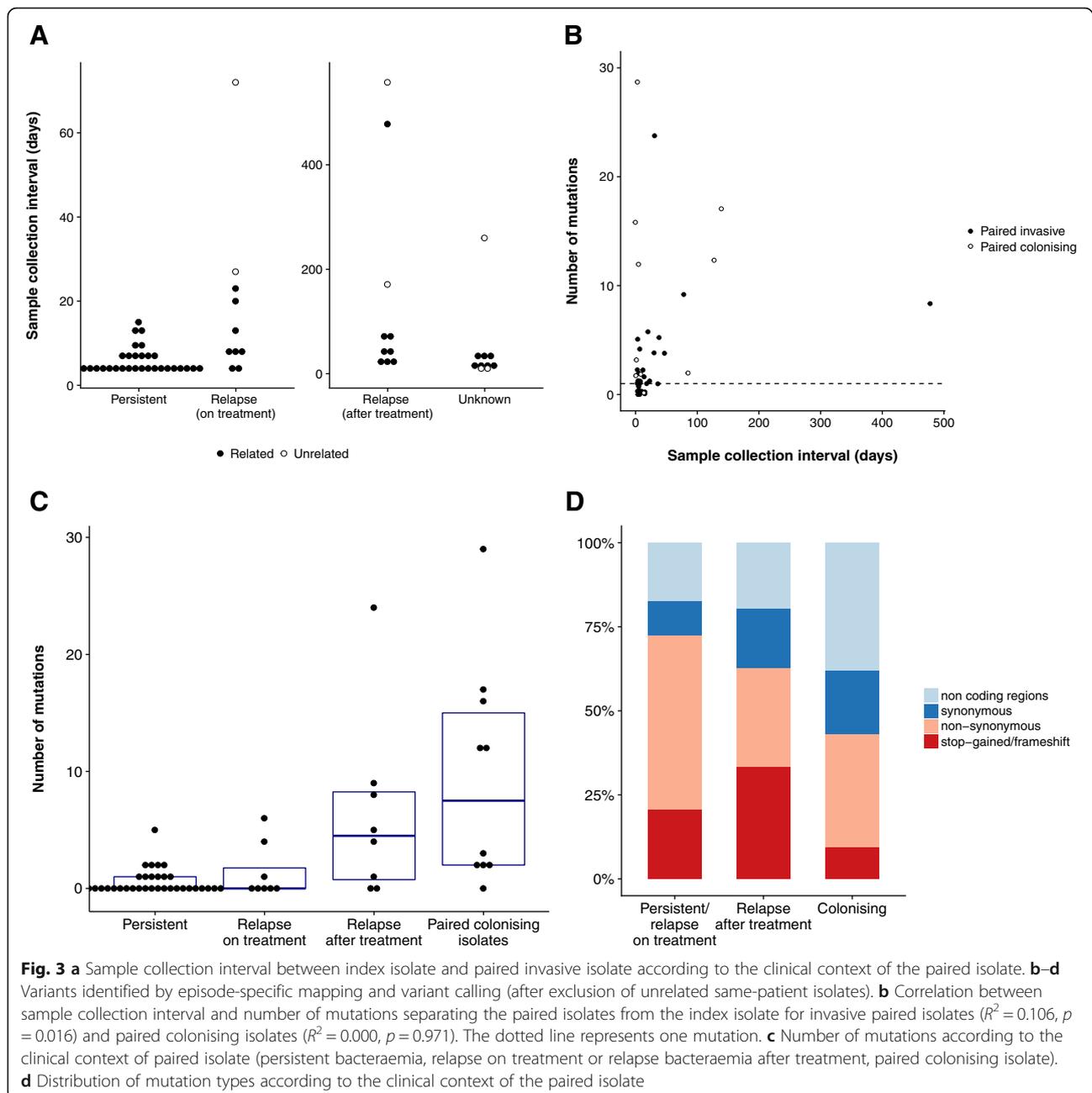
Colonising isolates and late relapses have a distinctive molecular signature

While the rate of mutation in *S. aureus* may be dependent on the genetic background [49], it is unknown whether

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 7 of 17

Fig. 2 a Maximum-likelihood tree of 130 isolates from 57 patients with *Staphylococcus aureus* bacteraemia, rooted using *Staphylococcus argenteus* as outgroup. Patient-specific shape-colour combinations annotate branch tips. White circles indicate nodes with ≥ 95% ultrafast support and ≥ 80% SH-like approximate likelihood ratio test support. Frequency distribution of pairwise single-nucleotide polymorphism (SNP) distance between isolates from the same patient using the common reference *S. aureus* TW20 (**b**) and the *de novo* assembly of the index isolate (**c**)

evolution rates are different during invasive infection, where host immune response and antibiotic treatment exert a strong selective pressure. We therefore explored associations between mutation counts and clinical, phenotypic, and genetic characteristics of the paired isolates. No association was found between mutation count and MRSA status or clonal complex. Interestingly, while there was a weak correlation between length of the collection interval of invasive isolates and mutation count (Fig. 3b), the association was not linear, with an increase in

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 8 of 17

Fig. 3 a Sample collection interval between index isolate and paired invasive isolate according to the clinical context of the paired isolate. **b–d** Variants identified by episode-specific mapping and variant calling (after exclusion of unrelated same-patient isolates). **b** Correlation between sample collection interval and number of mutations separating the paired isolates from the index isolate for invasive paired isolates ($R^2 = 0.106$, $p = 0.016$) and paired colonising isolates ($R^2 = 0.000$, $p = 0.971$). The dotted line represents one mutation. **c** Number of mutations according to the clinical context of paired isolate (persistent bacteraemia, relapse on treatment or relapse bacteraemia after treatment, paired colonising isolate). **d** Distribution of mutation types according to the clinical context of the paired isolate

mutation counts when the collection interval exceeded 15 days (Additional file 2: Figure S4). Since 15 days is the usual duration of treatment of uncomplicated SAB [50], this suggests that genetic diversity was higher when the paired isolate was collected after treatment. Consistent with this observation, we found a significantly higher number of mutations in paired invasive isolates from relapses after completion of anti-staphylococcal treatment (median 4.5 mutations per isolate) as compared to isolates from persistent bacteraemia or relapses on treatment (median 0 mutations) (Fig. 3c). In terms of genetic diversity, isolates from relapses after treatment were as genetically

diverse as paired colonising isolates compared to index isolates (Fig. 3c), indicating that they might represent re-infection with a closely related strain (in other words, a new invasive event from the colonising compartment) rather than the result of a persistent invasive focus. Episode-specific phylogenies for two patients for whom multiple isolates were available (patients 27 and 38) confirm this pattern (Additional file 2: Figure S5).

A similar pattern was discovered when we analysed the predicted mutation effects on the encoded proteins. The proportion of non-silent mutations (either nonsynonymous or stop-gained or frameshift) decreased

Giulieri et al. Genome Medicine (2018) 10:65

Page 9 of 17

progressively from 72% in invasive isolates from persistent bacteraemia or relapse on treatment to 63% in isolates from relapse after treatment to 43% in colonising isolates (Fig. 3d). The high proportion of non-silent mutations (66% vs. 43%, p = 0.002) indicates that the invasive compartment may be under stronger positive selection compared to the "colonising compartment". On the other hand, mutations found in late relapses might arise in the colonising compartment rather than during invasive infection.

Adaptation pathways are episode-specific

To identify possible convergence of mutation pathways among the 82 variants associated with predicted change in protein sequences, we applied protein sequences clustering using CD-HIT (Additional file 4: Table S3). Overall, mutation pathways were highly diverse and episode-specific. The only protein-coding gene that was mutated in more than one episode was the accessory gene regulator component agrA, with a nonsynonymous SNP in a paired invasive isolate (T88M) and a frameshift in a colonising isolate (at position 127). No convergence was observed among mutations arising in intergenic regions.

Given the weak convergence among mutated genes, we attempted to identify common pathways of within-host diversity by categorising the mutated proteins using the Clusters of Orhologous Groups (COG) database and performing an enrichment analysis using reference genome S. aureus TW20 as a comparator. Analysis of 17 categories did not show any significant enrichment. Nevertheless, genes related to cell wall and membrane biogenesis among paired invasive isolates reached the lowest p value (uncorrected p value 0.063, Additional file 2: Figure S6). Overall, different pathways were affected by mutations in invasive and colonising pairs, an observation that is consistent with distinctive selective pressures in the nasal and the blood compartment.

Mutations in pairs with observed phenotypic changes
Antibiotic resistance and growth rate

Within-host phenotypic adaptation might indicate diversifying selection under the selective pressure of antibiotics and the immune system [51]. Therefore, we identified pairs with changes in specific phenotypes between the index isolate and paired isolates. We selected invasive pairs, as they were associated with a stronger positive selection signature with a higher proportion of non-silent mutations. We performed a pairwise analysis of phenotypes related to commonly used antibiotics for treating invasive S. aureus infections that may have changed in response to selective pressure (vancomycin MIC and oxacillin MIC, growth rate) (Additional file 2: Figure S7). Significant changes in phenotypes were observed in a small number of episodes (increase in vancomycin MIC of ≥ 1 μg/ml, 3 episodes, sharp decrease in growth rate leading to small colony variants [SCV], 3 episodes). Details of these episodes are shown in Table 1.

Genetic changes underlying the SCV phenotype and secondary increase in vancomycin MIC were diverse, indicating that phenotypic convergence was not associated with genetic convergence (Table 2). Patient 3 presented with a ST45-MRSA bacteraemia, which was associated with a dialysis-catheter device and was treated with vancomycin for 14 days. She had recurrent bacteraemia 38 days after the first episode. The recurrent strain exhibited an increase in vancomycin MIC from 0.75 to 2 μg/ml and a SCV phenotype. We identified four mutations arising in the relapsing strain: a non-synonymous SNP in the rpoB gene leading to an arginine-histidin substitution in position 503 (R503H), a deletion in position 283 of the rplV gene (ribosomal protein 22), a deletion in position 66 of the rplD gene (ribosomal protein 4), and a deletion leading to a truncation of gene ptsG. The rpoB R503H and ribosomal protein mutations have been previously described in in vitro selected vancomycin-intermediate mutants [52, 53], but never in clinical isolates. The rpoB R503H mutation is not associated with rifampicin resistance, consistent with the lack of exposure to rifampicin in this case. Patient 5 was treated with vancomycin for 3 days and flucloxacillin for 13 days for a ST15-MSSA catheter-related bacteraemia and experienced relapse at day 18. The relapsing strain had a SCV phenotype and vancomycin MIC increased from 1.5 to 2 μg/ml. It had a nonsynonymous SNP in the serine/threonine phosphatase (stp) gene leading to a N137D mutation. stp mutations have been identified previously in persistent SAB with secondary development of vancomycin-intermediate S. aureus (VISA) [54, 55]. Finally, patient 30 had relapsing ST5-MRSA bacteraemia after 14 days of vancomycin (combined with rifampicin and ciprofloxacin). Median doubling time of the relapsing strain increased from 42 to 83 min, while vancomycin MIC increased from 1.5 to 2 μg/ml. Mutation analysis identified a deletion in a protein whose function could not be predicted and a non-synonymous SNP (mutation A60D) in the ywlC gene, which encodes a translational factor (threonylcarbamoyl-AMP synthase) [56]. This gene has never been linked to vancomycin resistance or growth rate. However, a ywlC ortholog has been shown to be essential in E. coli [56]; thus, it is possible that point mutations impair S. aureus growth.

In two episodes with an increase in vancomycin MIC by at least 1 μg/ml, no mutation separated the index isolate from the paired isolate, suggesting that other genetic changes may have occurred (see below).

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 10 of 17

Table 1 Summary of episodes with phenotypic changes

Patient code	Type of isolate	Collection interval (days)	Context of paired isolate	Days of treatment	Principal antibiotic	*mecA*	Van MIC (µg/ml)	Median doubling time	Mutations (SNP/indels)	Non-silent mutations
Small-colony variant and vancomycin MIC increase										
P_03	Index	–	–	–	–	Pos	0.75	34.5	–	0
P_03	Paired invasive	38	Relapse (after treatment)	14	Van	Pos	2	53.9	4 (1/3)	4
P_05	Index	–	–	–	–	Neg	1.5	38.4	–	0
P_05	Paired invasive	18	Relapse (after treatment)	12	Flx	Neg	2	50.6	1 (1/0)	1
P_30	Index	–	–	–	–	Pos	1.5	42.1	–	0
P_30	Paired invasive	14	Persistent	14	Van	Pos	2	83.5	2 (1/1)	2
Vancomycin MIC increase										
P_42	Index	–	–	–	–	Pos	2	44.7	–	0
P_42	Paired invasive	9	Persistent	8	Van	Pos	3	46.5	0 (0/0)	0
P_46	Index	–	–	–	–	Pos	2	50.6	–	0
P_46	Paired invasive	23	Relapse (on treatment)	21	Van	Pos	3	48.9	0 (0/0)	0

Van vancomycin, *Flx* flucloxacillin, *MIC* minimum inhibitory concentration, *SNP* single nucleotide polymorphism

Cytotoxicity

Recently, it has been shown that within-host evolution from colonising to invasive *S. aureus* can be associated with a dramatic decrease in cytotoxicity [57]; however, it is unknown whether a similar trend can be observed during persistent infection. To assess evolution of cytotoxicity during SAB, we tested a subset of 21 episodes that were considered more likely to be associated with changes based on phenotypic characteristics (i.e. longer duration of bacteraemia, relapse on anti-staphylococcal treatment, small colony phenotype or secondary increase in vancomycin MIC) or because of mutations in the *agrA* gene.

Similar to the other pairwise phenotypic tests, cytotoxicity remained unchanged between index isolate and paired isolate, with the dramatic exception of one paired invasive isolate with *agrA* mutation T88M, which was associated with a marked reduction in THP-1 cell lysis (from 56% non-viable cells to 11%) as compared to the index isolate (Additional file 2: Figure S7).

Table 2 Mutations in episodes with phenotypic changes

Patient code	Gene	Type	Mutation	Product	Category	Significance
P_03	*ptsG*	del	G306fs (stop at residue 341/682)	PTS system glucose-specific EIICBA component	Carbohydrate transport and metabolism	Three-component glucose transporter with phosphorylation activity [73]. *ptsG* deletion associated to resistance to glycosylated bacteriocins [74]
P_03	*rpoB*	snp	R503H	DNA-directed RNA polymerase subunit beta	Transcription	R503H associated with VISA phenotype in vitro. No rifampicin resistance [52]
P_03	*rplV*	complex	AIN95GR	50S ribosomal protein L22	Translation, ribosomal structure and biogenesis	Association with slow growth in in vitro selected VISA harbouring *rpoB* A621E [53]
P_03	*rplD*	del	KG68del	50S ribosomal protein L4	Translation, ribosomal structure and biogenesis	Mutations at positions 68 and 69 associated with linezolid resistance [75] and macrolide resistance [76]
P_05	*stp*	snp	N137D	Serine/threonine phosphatase stp	Signal transduction mechanisms	VISA phenotype in clinical strains; confirmed by mutagenesis [54, 55]
P_30	–	del	HVC139R	hypothetical protein	Function unknown	
P_30	*ywlC*	snp	A60D	Threonylcarbamoyl-AMP synthase	Translation, ribosomal structure and biogenesis	Required for the attachment of a threonylcarbamoyl group to ANN-decoding tRNA [56]

PTS phosphotransferase system, *VISA* vancomycin-intermediate *Staphylococcus aureus*, *AMP* adenosine monophosphate

Giulieri et al. Genome Medicine (2018) 10:65

Page 11 of 17

Chromosome structural variants

The potential significance of chromosome structural variants in *S. aureus* resistance and adaptation has been recently highlighted [13, 58]. However, when using only partially assembled genomes or read-mapping, the characterisation of structural variants is much more challenging than SNP calling and these types of changes are often overlooked. Using an approach combining read coverage arithmetic, read filtering, and annotation of split reads (Fig. 1e, f), we detected 21 unique structural variants within 15 SAB episodes: two plasmid losses, five large deletions (ranging from 261 to 15,622 bp), one recombination, and 13 insertions (summarised in Table 3). Beside the two instances of plasmid loss, which may have occurred after isolate collection/storage, two large structural changes were particularly interesting. A recombination of prophage Sa phi3 encoding the immune escape cluster (IEC) was identified when comparing the index blood isolate of patient 19 with its paired colonising isolate, based on accumulation of 192 mutations and discovery of a 6643-bp deletion on the index isolate. As a consequence, the invasive isolate carried the staphylokinase gene (*sak*) and the staphylococcal complement inhibitor gene (*scn*), while the colonising isolate carried the complete IEC including *sak*, *scn*, and the chemotaxis inhibitory protein (*chp*). Since there were no enterotoxin genes, this constellation can be classified as IEC types E and B, respectively, according to the classification proposed by van Wamel et al. [59]. In patient 21, we observed a deletion of pathogenicity island SaPi2 in the paired invasive isolate that was collected upon relapse of bacteraemia 65 days after the first episode. To complement the exploration of chromosome structural variants, we performed a pan-genome analysis, which confirmed major changes already identified (two plasmid losses, phage recombination, and SaPi2 deletion) and allowed us to identify a previously undetected loss of an ICE*6013* element.

The most prevalent structural change in our cohort was the insertion of IS*256* elements. We observed 13 unique IS*256* insertions in 8 episodes. Interestingly, the number of insertions did not correlate with the number of mutations, and up to three new IS*256* insertions were found in paired invasive isolates with no mutations relative to index (Additional file 2: Figure S8).

Intriguingly, all strains with new insertions belonged to ST239 or to a closely related single-locus variant of ST239. A BLAST search of the 1324-bp-long IS*256* sequence among all available *S. aureus* complete genomes and the draft assemblies of the 130 isolates included in our study confirmed that IS256 is highly disseminated in ST239 and restricted to a few other sequence types (Additional file 2: Figure S9).

We mapped split reads that were unique within single episodes on a single ST239 reference genome (*S. aureus*

TW20) and found that there were hotspots for these new IS*256* insertions on the chromosome (Fig. 4). One of these hotspots was the genomic island niSa beta with unique new insertions in four different patients, two of them around the lantibiotics operon. Moreover, we discovered that the paired invasive isolate from patient 46 had three new IS*256* insertions including one 150 bp upstream of the *walKR* operon. This finding was relevant because the isolate showed an increase in vancomycin MIC from 2 to 3 μg/ml as compared to the index isolate but no point mutations were found (see above). Notably, IS*256* insertion and tempering of WalKR activity has been previously shown to cause VISA phenotype in vitro but has never been described during human infection [58].

Discussion

This large-scale comparative genomics study of patients with persistent or recurrent SAB provides the first comprehensive overview of within-host *S. aureus* diversity associated with bacteraemia. By compiling a curated inventory of mutational and structural within-host variants across different genetic backgrounds and manifestations of SAB, we show that invasive isolate pairs have a specific molecular signature (denoted by limited diversity and a high proportion of non-silent mutations) and that structural variation and especially insertion of IS*256* elements enhances genetic diversity during human infection. With the notable exception of *agrA*, which was mutated in one invasive pair and one invasive-colonising pair, there was no convergence at the gene level among mutations and indels, indicating that pathways of adaption are episode-specific, even when we found common phenotypic changes within pairs. Loss of *agr* function within the host has been previously described both among colonising isolates [5] and during invasive infection [8, 60]. Enrichment for *agrA* mutations was also observed in a study of 105 colonising-invasive pairs [7]. While one of the two mutations was associated with the only significant reduction in cytotoxicity observed in our cohort, none led to an increase in vancomycin MIC, despite the known link between *agr* dysfunction and vancomycin resistance [61].

When considering bacterial within-host diversity in invasive *S. aureus* infections, it is important to keep a "pathogen-centric" perspective and consider a model consisting of two "compartments", i.e. the colonising compartment (anterior nares, or other mucosal areas) and the invasive compartment (blood and tissue/organs of primary or metastatic *S. aureus* infection). Bacteria in the colonising compartment are subjected to evolution pressures (competition of the nasal microbiota, some immune system control and intermittent antibiotic exposure at low concentration) but also to purifying

Giulieri et al. Genome Medicine (2018) 10:65

Page 12 of 17

Table 3 Overview of chromosome structural variants

Patient code	Type of variant	Reference (position)	Annotation	Isolate with variant
P_27	Plasmid loss	S. aureus TW20	29,585 bp-plasmid containing the chlorhexidine tolerance determinant qacA and cadmium resistance genes	Index, paired invasive, one paired colonising (second paired colonising like reference)
P_30	Plasmid loss	S. aureus SA564	27,272 bp-plasmid containing the beta-lactamase operon	Index (paired invasive like reference)
P_05	Deletion	S. aureus NCTC_8325 (2188042)	861 bp-deletion (fructose bisphosphate aldolase)	Paired invasive
P_12	Deletion	S. aureus NCTC_8325 (1502725)	261 bp-deletion (phage protein)	Index (paired invasive like reference)
P_14	Deletion	S. aureus FORC_001 (1953000)	597 bp-deletion (hypothetical protein)	Paired invasive
P_21	Deletion	S. aureus SA564 (798999)	15,622 bp deletion of pathogenicity island SaPi2	Paired invasive
P_54	Deletion	S. aureus AUS0325 (1191813)	495 bp-deletion (serine/threonine-protein phosphatase)	Index (paired colonising like reference)
P_19	Recombination	S. aureus CA-347 (1577530)	Phage Sa phi3 recombination	Index (paired colonising like reference)
P_22	Insertion	S. aureus TW20 (2243085)	IS256 insertion (prophage phiSPbeta-like)	Paired invasive
P_27	Insertion	S. aureus TW20 (173671)	IS256 insertion (putative DNA-binding protein)	Index and one paired invasive
P_27	Insertion	S. aureus TW20 (1223135)	IS256 (putative membrane protein)	Index, paired invasive (paired colonising like reference)
P_27	Insertion	S. aureus TW20 (1971406)	IS256 insertion (lantibiotic immunity protein, genomic island niSa beta)	Index, paired invasive, one paired colonising (second paired colonising like reference)
P_27	Insertion	S. aureus TW20 (173,671, 173,626)	IS256 insertion (putative DNA-binding protein)	Index and one paired invasive
P_34	Insertion	S. aureus TW20 (1619645)	IS256 insertion (hypothetical protein)	Index (paired invasive like reference)
P_39	Insertion	S. aureus TW20 (1955676)	IS256 insertion (genomic island niSa beta)	Index (paired invasive like reference)
P_46	Insertion	S. aureus TW20 (24644)	IS256 insertion (upstream of walKR operon)	Paired invasive
P_46	Insertion	S. aureus TW20 (528848)	IS256 insertion (exotoxin, genomic island niSa alpha)	Paired invasive
P_46	Insertion	S. aureus TW20 (2306594)	IS256 insertion (phage protein)	Paired invasive
P_47	Insertion	S. aureus str._JKD6008 (1914213)	IS256 insertion (genomic island niSa beta)	Paired invasive
P_47	Insertion	S. aureus str._JKD6008 (2227509)	IS256 insertion (P-ATPase superfamily P-type ATPase potassium (K+) transporter subunit A)	Paired invasive
P_49	Insertion	S. aureus str._JKD6008 (702665)	IS256 insertion (upstream of sodium/hydrogen exchanger family protein)	Paired invasive
P_52	Insertion	S. aureus TW20 (1955675)	IS256 insertion (genomic island niSa beta)	Paired colonising

Bp base pair

selection, since colonisation sites such as the nose are the natural ecological niche of S. aureus [51, 62]. By contrast, bacterial invading blood and tissue are subjected to a formidable selective pressure, including antibiotics at high concentration, host antimicrobial peptides, immune cells, and sequestration of nutrients (e.g. iron). This is supported by convergent evolution analysis at the gene ontology level. In this study, a non-significant enrichment for mutations in genes associated with cell wall and membrane metabolism was found in paired invasive isolates, while enrichment for genes associated with cell wall and

adhesion was described by Young et al. among colonising-invasive pairs [7].

Since the advent of WGS, studies addressing within-host diversity of S. aureus bacteraemia have mainly focused on genetic changes associated with secondary development of the VISA phenotype under vancomycin pressure [8, 9, 11, 13, 54, 63], or with secondary development of daptomycin resistance [10]. By selecting SAB episodes with phenotypic changes, this approach helps to distinguish evolution in the blood/tissue compartment from background diversity of the colonising S. aureus

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 13 of 17

Fig. 4 Location of IS*256* insertions differentiating CC239 isolates within the same patient. Insertions are mapped on the chromosome of the reference *S. aureus* TW20 and labelled with the patient ID. Two sites of interest are depicted in detail. The diagram of site 1 (position 24,645-27,443) shows an IS*256* insertion 150 bp upstream of the two-component regulator *walKR*, that was found in the paired invasive isolate but not in the index isolate of patient 46. Site 2 (position 1,950,453-1,984,290) is the genomic island niSa-beta, which appears to be a hotspot of new IS*256* insertions within same-patient isolates

population. Our study complements this previous work by providing for the first time a wider picture of within-host diversity in SAB in a diverse genetic background. By using phenotypic tests such as vancomycin susceptibility testing and growth curves, we showed that secondary changes were present in a small proportion of cases. However, even in this group with more evident features of positive evolution, we found a heterogeneity of mutations. This observation, together with the wide range of alleles described in the VISA literature [61], highlights the multiplicity of pathways by which *S. aureus* adapts to vancomycin pressure in vivo. Mutations identified in episodes without detected phenotypic changes were also very heterogenous, and thus we were not able to detect convergence in our dataset or identify mutational hotspots that were associated with persistence or recurrence. This clearly shows that we should be careful in drawing general conclusions on *S. aureus*

pathoadaptation from mutations identified in single clinical cases.

One striking finding was the combination of limited genetic variability and high frequency of non-silent mutations among invasive pairs. This is consistent with the "bottle neck" hypothesis of SAB, as shown in animal models, where only individual clones among the diverse colonising pool become invasive [64]. Despite the lack of identified molecular hotspots of persistence, this signature (low abundance of mutations and low fraction of "silent mutations") may help distinguish between reinfection with a close related colonising strain from relapse from a persistent infection focus. On the background of increased availability of WGS, within-host diversity data could be used not only to understand the pathogenesis of SAB and antibiotic resistance, but also to inform clinical management of persistence or recurrence.

Giulieri et al. Genome Medicine (2018) 10:65

Page 14 of 17

Mobile genetic elements (MGE) are key drivers of evolution in *S. aureus* [65]. Evolution experiments in vivo have shown an intense exchange of MGE in piglets co-colonised with different lineages of *S. aureus* [66], and there is evidence of phage recombination in studies of same-patient colonising isolates [5, 62]. In addition to MGE, we have recently illustrated that a large chromosome duplication mediated vancomycin resistance and immune evasion in a case of extremely protracted SAB [13]. However, we do not have an overview of structural variation and MGE movements during clinical invasive *S. aureus* infection. Therefore, we applied an episode-specific strategy to detect structural variation by carefully assessing read coverage and split reads within the pairs to identify unique structural changes that were confirmed by review of the assembly graphs. This mapping-based approach allowed us to reveal large changes occurring even in the absence of point mutations. Some of these modifications may have a relevant impact on the phenotype. For example, we observed the loss of pathogenicity island SaPI2 in a paired invasive isolate and a recombination of phage Sa phi3 (including the IEC) in an invasive-colonising pair.

While deletion and recombination were episode-specific without a discernible pattern, a striking finding was the remarkable high frequency of new IS256 insertions within the same episode among strains belonging to the dominant lineage ST239 (at least one insertion event in 8 out of 13 episodes) with the genomic island niSa beta as a hotspot of new insertions. This genomic island is enriched with IS256 that has been shown to engender chromosomal inversion in an ST8-IV MRSA [67]. The effect of the new IS256 insertions in genomic island niSa beta in our paired clinical isolates are uncertain, although insertions around the lantibiotic operon could be important in modulating the production of lantibiotics depending on whether *S. aureus* is in the colonising compartment (i.e. competing with other microbiota) or is invasive [68]. Up to three new insertions occurred in paired isolates even in the absence of point mutations, suggesting that IS256 is an efficient mechanism of genetic variability in the environment of invasive *S. aureus* infection, characterised by high selective pressure and reduced effective population size, as it is known that bacterial stress like antibiotic exposure activates insertion sequences [69, 70]. A paradigmatic example was the insertion of IS256 upstream of the *walKR* operon (in an isolate whose vancomycin MIC increased from 2 to 3 μg/ml), a mechanism that we and others have previously elucidated in in vitro selected VISA strains [58, 71]. Based on these latter studies and others [72] showing that IS256 is selected for in vitro, we hypothesise that the new insertions were mediated by a "copy-paste" mechanism rather than transfer from an external donor (within the colonising compartment); however, short-read data do not allow to assess IS256 diversity within the same genome.

Our study has some limitations. Because patients were recruited at detection of positive blood cultures for *S. aureus*, colonising *S. aureus* strains were available only for a very small proportion of patients. Therefore, we included invasive-colonising pairs as a comparator to invasive pairs, but our dataset prevents conclusions on molecular signatures on "invasiveness" of *S. aureus*. Furthermore, in our study, one colony per sample was sequenced. Recent work has exposed the diversity of colonising and invasive *S. aureus* strains by sequencing multiple colonies per sample (up to 12) [5, 6]. Data obtained from this "high resolution" approach can then be used to better infer within-host phylogenies and shed light into the pathogenesis of SAB. Because of this limitation and the fact that for most episodes only one additional invasive isolate was available, care should be taken in driving conclusions on within-host evolution during SAB, since the genetic changes we observed may only represent inherent diversity rather than an evolutionary process. Additionally, we tested only a limited array of phenotypes. Data on more complex phenotypes (e.g. immune evasion) might have furnished additional insights into the impact of the host immunity on *S. aureus* evolution within the blood or tissue compartment. Finally, since our analysis of structural variants is based on short read data, we may have missed structural changes that can be usually only detected by long-read sequencing, such as chromosomal inversion.

Conclusions

By applying comparative genomics to 57 episodes of SAB with sequential invasive *S. aureus* isolates or paired colonising isolates, we describe specific patterns of *S. aureus* diversity within the invasive compartment (in particular limited within-host diversity and strong positive selection signatures), that might be indicative of the adaptive changes under the combined pressure of antibiotics and host immunity. Furthermore, we highlight the crucial role of structural changes and in particular MGEs like insertion sequences in conferring genetic plasticity within the host. Data from this study will improve our understanding of the bacterial pathogenesis of SAB and contribute to defining a molecular signature of persistence/relapse that might be used for both biological research and infection management.

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 15 of 17

Additional files

Abbreviations
CC: Clonal complex; MIC: Minimum inhibitory concentration; MRSA: Methicillin-resistant *Staphylococcus aureus*; MSSA: Methicillin-susceptible *Staphylococcus aureus*; SAB: *Staphylococcus aureus* bacteraemia; SCV: Small-colony variant; SNP: Single-nucleotide polymorphism; ST: Sequence type; VISA: Vancomycin-intermediate *Staphylococcus aureus*; WGS: Whole-genome sequencing

Availability of data and materials
The datasets supporting the conclusions of this article are available in the European Nucleotide Archive under Bioproject PRJEB27932.

Authors' contributions
SGG and BPH designed and planned the study. SGG, SLB, NEH, and BPH supplied isolates, clinical data, and whole-genome sequencing. SGG, SLB, and NEH performed laboratory experiments. SGG, SLB, RG, TS, AGS, MS, RM, NEH, TPS, and BPH analysed the data. SGG and BPH drafted the manuscript. All authors reviewed and contributed to the final manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Human research ethics committee approval was obtained at each of the centres participating to the ANZCOSS and VANESSA cohorts (Victoria, Australia: Austin Health, Alfred Health, Southern Health, Barwon Health; New South Wales, Australia: Liverpool Hospital, Royal Prince Alfred Hospital, Westmead Hospital; Queensland, Australia: Princess Alexandra Hospital; West Australia: Royal Perth Hospital; Tasmania, Australia: Royal Hobart Hospital; Auckland, New Zealand: Auckland District Health Board). Written informed consent was obtained from participants to the prospectively collected VANESSA cohort. Consent from each patient for ANZCOSS was not required as the collection of data was considered to be a clinical audit. Research was performed in accordance with the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Microbiology and Immunology, The University of Melbourne at the Doherty Institute for Infection & Immunity, Melbourne, Australia. [2]Infectious Disease Department, Austin Health, Melbourne, Australia. [3]Infectious Diseases Service, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland. [4]Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne at The Doherty Institute of Infection and Immunity, Melbourne, Australia. [5]Melbourne Bioinformatics, The University of Melbourne, Melbourne, Australia. [6]School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK.

References
1. Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG Jr. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev. 2015;28:603–61.
2. Hawkins C, Huang J, Jin N, Noskin GA, Zembower TR, Bolon M. Persistent Staphylococcus aureus bacteremia: an analysis of risk factors and outcomes. Arch Intern Med. 2007;167:1861–7.
3. Xiong YQ, Fowler VG, 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;199:201–8.
4. Giulieri SG, Holmes NE, Stinear TP, Howden BP. Use of bacterial whole-genome sequencing to understand and improve the management of invasive Staphylococcus aureus infections. Expert Rev Anti-Infect Ther. 2016; 14:1023–36.
5. Paterson GK, Harrison EM, Murray GG, Welch JJ, Warland JH, Holden MT, Morgan FJ, Ba X, Koop G, Harris SR, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. Nat Commun. 2015;6:6560.
6. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, et al. Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. Proc Natl Acad Sci U S A. 2012;109:4550–5.
7. Young BC, Wu C-H, Gordon NC, Cole K, Price JR, Liu E, Sheppard AE, Perera S, Charlesworth J, Golubchik T, et al. Severe infections emerge from commensal bacteria by adaptive evolution. eLife. 2017;6:e30637.
8. Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, Tomasz A. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Proc Natl Acad Sci U S A. 2007;104:9451–6.
9. Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, et al. Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. PLoS Pathog. 2011;7:e1002359.
10. Peleg AY, Miyakis S, Ward DV, Earl AM, Rubio A, Cameron DR, Pillai S, Moellering RC Jr, Eliopoulos GM. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of Staphylococcus aureus. PLoS One. 2012;7:e28316.
11. Gao W, Chua K, Davies JK, Newton HJ, Seemann T, Harrison PF, Holmes NE, Rhee HW, Hong JI, Hartland EL, et al. Two novel point mutations in clinical

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 16 of 17

Staphylococcus aureus reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. PLoS Pathog. 2010;6: e1000944.

12. Kim NH, Kang YM, Han WD, Park KU, Park KH, Yoo JI, Lee DG, Park C, Song KH, Kim ES, et al. Small-Colony variants in persistent and recurrent Staphylococcus aureus bacteremia. Microb Drug Resist. 2016;22:538–44.

13. Gao W, Monk IR, Tobias NJ, Gladman SL, Seemann T, Stinear TP, Howden BP. Large tandem chromosome expansions facilitate niche adaptation during persistent infection with drug-resistant Staphylococcus aureus. Microbial Genomics. 2015;1:e000026.

14. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017;545:446–51.

15. Holmes NE, Turnidge JD, Munckhof WJ, Robinson JO, Korman TM, O'Sullivan MV, Anderson TL, Roberts SA, Gao W, Christiansen KJ, et al. Antibiotic choice may not explain poorer outcomes in patients with Staphylococcus aureus bacteremia and high vancomycin minimum inhibitory concentrations. J Infect Dis. 2011;204:340–7.

16. Turnidge JD, Kotsanas D, Munckhof W, Roberts S, Bennett CM, Nimmo GR, Coombs GW, Murray RJ, Howden B, Johnson PD, et al. Staphylococcus aureus bacteraemia: a major cause of mortality in Australia and New Zealand. Med J Aust. 2009;191:368–73.

17. Holmes NE, Turnidge JD, Munckhof WJ, Robinson JO, Korman TM, O'Sullivan MV, Anderson TL, Roberts SA, Warren SJ, Coombs GW, et al. Genetic and molecular predictors of high vancomycin MIC in Staphylococcus aureus bacteremia isolates. J Clin Microbiol. 2014;52:3384–93.

18. Holmes NE, Turnidge JD, Munckhof WJ, Robinson JO, Korman TM, O'Sullivan MV, Anderson TL, Roberts SA, Warren SJ, Gao W, et al. Vancomycin AUC/MIC ratio and 30-day mortality in patients with Staphylococcus aureus bacteremia. Antimicrob Agents Chemother. 2013;57:1654–63.

19. Holmes NE, Turnidge JD, Munckhof WJ, Robinson JO, Korman TM, O'Sullivan MV, Anderson TL, Roberts SA, Warren SJ, Gao W, et al. Vancomycin minimum inhibitory concentration, host comorbidities and mortality in Staphylococcus aureus bacteraemia. Clin Microbiol Infect. 2013;19:1163–8.

20. Holmes NE, Robinson JO, van Hal SJ, Munckhof WJ, Athan E, Korman TM, Cheng AC, Turnidge JD, Johnson PDR, Howden BP, Vanessa study group obotASfIDCRN. Morbidity from in-hospital complications is greater than treatment failure in patients with Staphylococcus aureus bacteraemia. BMC Infect Dis. 2018;18:107.

21. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.

22. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46.

23. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4.

24. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004;186: 1518–30.

25. Holden MT, Lindsay JA, Corton C, Quail MA, Cockfield JD, Pathak S, Batra R, Parkhill J, Bentley SD, Edgeworth JD. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW). J Bacteriol. 2010;192:888–92.

26. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

27. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

28. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95.

29. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.

30. Tong SY, Schaumburg F, Ellington MJ, Corander J, Pichon B, Leendertz F, Bentley SD, Parkhill J, Holt DC, Peters G, Giffard PM. Novel staphylococcal species that form part of a Staphylococcus aureus-related complex: the

non-pigmented Staphylococcus argenteus sp. nov. and the non-human primate-associated Staphylococcus schweitzeri sp. nov. Int J Syst Evol Microbiol. 2015;65:15–22.

31. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90.

32. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36.

33. Collins J, Buckling A, Massey RC. Identification of factors contributing to T-cell toxicity of Staphylococcus aureus clinical isolates. J Clin Microbiol. 2008; 46:2112–4.

34. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, Williams P, Endres JL, Bayles KW, Fey PD, et al. Predicting the virulence of MRSA from its genome sequence. Genome Res. 2014;24:839–49.

35. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The sequence alignment/ map format and SAMtools. Bioinformatics. 2009;25:2078–9.

37. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944.

38. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

39. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25:1422–3.

40. Thorpe HA, Bayliss SC, Sheppard SK, Feil EJ. Piggy: a rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria. Gigascience. 2018;7: 1–1.

41. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

42. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.

43. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.

44. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.

45. Gagneur J, Neudecker A. cellGrowth: fitting cell population growth models. 1.18.0 ed; 2012. R package

46. Fowler VG Jr, Kong LK, Corey GR, Gottlieb GS, McClelland RS, Sexton DJ, Gesty-Palmer D, Harrell LJ. Recurrent Staphylococcus aureus bacteremia: pulsed-field gel electrophoresis findings in 29 patients. J Infect Dis. 1999;179:1157–61.

47. Briskine RV, Shimizu KK. Positional bias in variant calls against draft reference assemblies. BMC Genomics. 2017;18:263.

48. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, Driebe EM, Drees KP, Hicks ND, Williamson CHD, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microbial Genomics. 2016;2:e000074.

49. Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. Genome-scale rates of evolutionary change in bacteria. Microbial Genomics. 2016;2:e000094.

50. Thwaites GE, Edgeworth JD, Gkrania-Klotsas E, Kirby A, Tilley R, Torok ME, Walker S, Wertheim HF, Wilson P, Llewelyn MJ, Group UKCIR. Clinical management of Staphylococcus aureus bacteraemia. Lancet Infect Dis. 2011;11:208–22.

51. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016;14:150–62.

52. Matsuo M, Hishinuma T, Katayama Y, Cui L, Kapi M, Hiramatsu K. Mutation of RNA polymerase beta subunit (rpoB) promotes hVISA-to-VISA phenotypic conversion of strain Mu3. Antimicrob Agents Chemother. 2011;55:4188–95.

53. Cui L, Isii T, Fukuda M, Ochiai T, Neoh HM, Camargo IL, Watanabe Y, Shoji M, Hishinuma T, Hiramatsu K. An RpoB mutation confers dual heteroresistance to daptomycin and vancomycin in Staphylococcus aureus. Antimicrob Agents Chemother. 2010;54:5222–33.

54. Passalacqua KD, Satola SW, Crispell EK, Read TD. A mutation in the PP2C phosphatase gene in a Staphylococcus aureus USA300 clinical isolate with reduced susceptibility to vancomycin and daptomycin. Antimicrob Agents Chemother. 2012;56:5212–23.

Giulieri *et al. Genome Medicine* (2018) 10:65

Page 17 of 17

55. Cameron DR, Ward DV, Kostoulias X, Howden BP, Moellering RC Jr, Eliopoulos GM, Peleg AY. Serine/threonine phosphatase Stp1 contributes to reduced susceptibility to vancomycin and virulence in Staphylococcus aureus. J Infect Dis. 2012;205:1677–87.

56. El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crécy-Lagard V. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. Nucleic Acids Res. 2009;37:2894–909.

57. Laabei M, Uhlemann AC, Lowy FD, Austin ED, Yokoyama M, Ouadi K, Feil E, Thorpe HA, Williams B, Perkins M, et al. Evolutionary trade-offs underlie the multi-faceted virulence of Staphylococcus aureus. PLoS Biol. 2015;13: e1002229.

58. McEvoy CR, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP. Decreased vancomycin susceptibility in Staphylococcus aureus caused by IS256 tempering of WalKR expression. Antimicrob Agents Chemother. 2013;57:3240–9.

59. van Wamel WJ, Rooijakkers SH, Ruyken M, van Kessel KP, van Strijp JA. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages. J Bacteriol. 2006;188:1310–5.

60. Sakoulas G, Eliopoulos GM, Moellering RC Jr, Wennersten C, Venkataraman L, Novick RP, Gold HS. Accessory gene regulator (agr) locus in geographically diverse Staphylococcus aureus isolates with reduced susceptibility to vancomycin. Antimicrob Agents Chemother. 2002;46:1492–502.

61. Howden BP, Davies JK, Johnson PD, Stinear TP, Grayson ML. Reduced vancomycin susceptibility in Staphylococcus aureus, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. Clin Microbiol Rev. 2010;23:99–139.

62. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, Fung R, Godwin H, Knox K, Votintseva A, et al. Within-host evolution of Staphylococcus aureus during asymptomatic carriage. PLoS One. 2013;8: e61319.

63. Rishishwar L, Kraft CS, Jordan IK. Population genomics of reduced vancomycin susceptibility in Staphylococcus aureus. mSphere. 2016;1:e00094-16.

64. McVicker G, Prajsnar TK, Williams A, Wagner NL, Boots M, Renshaw SA, Foster SJ. Clonal expansion during Staphylococcus aureus infection dynamics reveals the effect of antibiotic intervention. PLoS Pathog. 2014;10: e1003959.

65. Malachowa N, DeLeo FR. Mobile genetic elements of Staphylococcus aureus. Cell Mol Life Sci. 2010;67:3057–71.

66. McCarthy AJ, Loeffler A, Witney AA, Gould KA, Lloyd DH, Lindsay JA. Extensive horizontal gene transfer during Staphylococcus aureus co-colonization in vivo. Genome Biol Evol. 2014;6:2697–708.

67. Wan TW, Khokhlova OE, Iwao Y, Higuchi W, Hung WC, Reva IV, Singur OA, Gostev VV, Sidorenko SV, Peryanova OV, et al. Complete circular genome sequence of successful ST8/SCCmecIV community-associated methicillin-resistant Staphylococcus aureus (OC8) in Russia: one-megabase genomic inversion, IS256's spread, and evolution of Russia ST8-IV. PLoS One. 2016;11: e0164168.

68. Daly KM, Upton M, Sandiford SK, Draper LA, Wescombe PA, Jack RW, O'Connor PM, Rossney A, Götz F, Hill C, et al. Production of the Bsa Lantibiotic by community-acquired Staphylococcus aureus strains. J Bacteriol. 2010;192:1131–42.

69. Schreiber F, Szekat C, Josten M, Sahl HG, Bierbaum G. Antibiotic-induced autoactivation of IS256 in Staphylococcus aureus. Antimicrob Agents Chemother. 2013;57:6381–4.

70. Nagy Z, Chandler M. Regulation of transposition in bacteria. Res Microbiol. 2004;155:387–98.

71. Jansen A, Turck M, Szekat C, Nagel M, Clever I, Bierbaum G. Role of insertion elements and yycFG in the development of decreased susceptibility to vancomycin in Staphylococcus aureus. Int J Med Microbiol. 2007;297:205–15.

72. Kleinert F, Kallies R, Hort M, Zweynert A, Szekat C, Nagel M, Bierbaum G. Influence of IS256 on genome variability and formation of small-colony variants in Staphylococcus aureus. Antimicrob Agents Chemother. 2017;61: e00144–17.

73. Christiansen I, Hengstenberg W. Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system—two highly similar glucose permeases in Staphylococcus carnosus with different glucoside specificity: protein engineering in vivo? Microbiology. 1999;145:2881–9.

74. Garcia De Gonzalo CV, Denham EL, Mars RAT, Stülke J, van der Donk WA, van Dijl JM. The phosphoenolpyruvate:sugar phosphotransferase system is involved in sensitivity to the glucosylated bacteriocin sublancin. Antimicrob Agents Chemother. 2015;59:6844–54.

75. Locke JB, Hilgers M, Shaw KJ. Novel ribosomal mutations in Staphylococcus aureus strains identified through selection with the Oxazolidinones linezolid and Torezolid (TR-700). Antimicrob Agents Chemother. 2009;53:5265–74.

76. Prunier A-L, Malbruny B, Laurans M, Brouard J, Duhamel J-F, Leclercq R. High rate of macrolide resistance in Staphylococcus aureus strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. J Infect Dis. 2003;187:1709–16.