Niche-specific genome degradation and convergent evolution shaping *Staphylococcus aureus* adaptation during severe infections

Stefano G Giulieri[1,2,3], Romain Guérillot[1], Sebastian Duchene[1], Abderrahman Hachani[1], Diane Daniel[1,4], Torsten Seemann[4], Joshua S Davis[5,6], Steven YC Tong[6,7], Bernadette C Young[8], Daniel J Wilson[9], Timothy P Stinear[1]*, Benjamin P Howden[1,2,4]*

[1]Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia; [2]Department of Infectious Diseases, Austin Health, Heidelberg, Australia; [3]Victorian Infectious Diseases Service, Royal Melbourne Hospital, Melbourne, Australia; [4]Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne at the Doherty Institute for Infection and Immunity, Melbourne, Australia; [5]Department of Infectious Diseases, John Hunter Hospital, Newcastle, New South Wales, Australia; [6]Menzies School of Health Research, Charles Darwin University, Casuarina, Northern Territory, Australia; [7]Victorian Infectious Disease Service, Royal Melbourne Hospital, and University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia; [8]Nuffield Department of medicine, Oxford, United Kingdom; [9]Big Data Institute, Nuffield Department of Population Health, Li Ka Shing Centre for Health Information and Discovery, Old Road Campus, University of Oxford, Oxford, United Kingdom

*For correspondence:
tstinear@unimelb.edu.au (TPS);
bhowden@unimelb.edu.au (BPH)

Abstract During severe infections, *Staphylococcus aureus* moves from its colonising sites to blood and tissues and is exposed to new selective pressures, thus, potentially driving adaptive evolution. Previous studies have shown the key role of the *agr* locus in *S. aureus* pathoadaptation; however, a more comprehensive characterisation of genetic signatures of bacterial adaptation may enable prediction of clinical outcomes and reveal new targets for treatment and prevention of these infections. Here, we measured adaptation using within-host evolution analysis of 2590 *S. aureus* genomes from 396 independent episodes of infection. By capturing a comprehensive repertoire of single nucleotide and structural genome variations, we found evidence of a distinctive evolutionary pattern within the infecting populations compared to colonising bacteria. These invasive strains had up to 20-fold enrichments for genome degradation signatures and displayed significantly convergent mutations in a distinctive set of genes, linked to antibiotic response and pathogenesis. In addition to *agr*-mediated adaptation, we identified non-canonical, genome-wide significant loci including *sucA-sucB* and *stp1*. The prevalence of adaptive changes increased with infection extent, emphasising the clinical significance of these signatures. These findings provide a high-resolution picture of the molecular changes when *S. aureus* transitions from colonisation to severe infection and may inform correlation of infection outcomes with adaptation signatures.

Editor's evaluation

This study offers a comprehensive examination of *Staphylococcus aureus* evolution during infection. It provides a useful analysis of select genetic signatures during bacterial adaptation. A combination of multiple layers of genome annotation and point mutation variant detection compellingly supports the correlation of infection outcomes with adaptation signatures in *S. aureus*.

Introduction

While *Staphylococcus aureus* is one of the most important human pathogens (**Tong et al., 2015**), its common interaction with the human host is colonisation, usually of the anterior nares (**Wertheim et al., 2005**). Comparatively, severe, life-threatening infections such as bacteraemia or osteomyelitis occur very rarely. This suggests that at the macro-evolutionary level, *S. aureus* is primarily adapted to its natural ecological niche (the nasal cavity) and to specific selective pressures arising in this environment, such as competition with the resident microbiota (**Krismer et al., 2017**). By contrast, during invasive infection, a new fitness trade-off needs to be achieved to adjust to environmental challenges that include innate and acquired immune responses (**Proctor, 2019**), high-dose antibiotics (**Kuehl et al., 2020**), and nutrient starvation (**Hood and Skaar, 2012**). These trade-offs could occur across three potentially distinctive dynamics of micro-evolution during colonisation and infection (within the colonising population, from colonising to invasive, and within the invasive population), leading to nose-adapted, early infection-adapted, and late infection-adapted strains. Identifying infection-adapted strains might assist precision medicine strategies for infection prevention and management and refine the understanding of *S. aureus* pathogenesis versatility, as mutational footprints of selection mirror functions that are critically important for bacterial survival during invasion.

Emerging genomic approaches for analysis of within-host evolution are among the most powerful means to study bacterial host adaptation (**Marvig et al., 2015**; **Kennemann et al., 2011**; **Lieberman et al., 2016**). Studies have shown the remarkable diversity and evolution of colonising populations of *Streptococcus pneumoniae* (**Chaguza et al., 2020**) and *S. aureus* (**Paterson et al., 2015**). In *S. aureus*, *Enterococcus faecalis* and *Enterococcus faecium* it has been shown that transition from colonisation to invasion favours strains with specific adaptive signatures (**Young et al., 2017**; **Van Tyne et al., 2019**; **Chilambi et al., 2020**), while evidence of niche adaptation was noted in a within-host study of bacterial meningitis due to *S. pneumoniae* (**Lees et al., 2017b**). Furthermore, phenotypic and genomic adaptation (often in response to antibiotic pressure) has been investigated during selected episodes of persistent invasive infections due to *S. aureus* (**Giulieri et al., 2020**; **Howden et al., 2006**; **Giulieri et al., 2018**), *Pseudomonas aeruginosa* (**Wheatley et al., 2021**), *Salmonella enterica* (**Klemm et al., 2016**), and *Mycobacterium tuberculosis* (**Vargas et al., 2021**). To increase power, bacterial within-host evolution studies have leveraged on large collections of paired samples coupled with statistical models of genome-wide mutation rates (**Marvig et al., 2015**; **Lees et al., 2017b**; **Gatt and Margalit, 2021**) and extended the analysis to include chromosomal structural variation (**Giulieri et al., 2018**; **Lees et al., 2017a**; **Gao et al., 2015**) as well as intergenic mutations (**Lees et al., 2017a**; **Khademi et al., 2019**).

Convergent evolution among separated (independent) episode of colonisation or infection is a key indication of adaptation in evolution analyses. However, with the notable exception of one study of cystic fibrosis (**Long et al., 2020**), the convergence has generally been weak in within-host studies of *S. aureus* infections, with no convergence at all (**Giulieri et al., 2018**) or significant enrichment limited to the *S. aureus* master regulator *agr* (**Young et al., 2017**). We hypothesised that in addition to the small sample size, the extended range of bacterial functions potentially under selective pressure (each function being potentially targeted by diverse pathoadaptive mutations) has hampered the identification of important adaptation mechanisms. To overcome the limitations of studies to date, we have pooled all publicly available *S. aureus* within-host evolution studies, and complemented this with a new dataset from a recent *S. aureus* clinical trial (**Tong et al., 2020**), in a single large-scale analysis. Rather than focussing on point mutations and small insertions/deletions alone, we leveraged multiple layers of genome annotation (encompassing coding regions, operons, intergenic regions, and functional categories) and included chromosome structural variants to compile a comprehensive catalogue of bacterial genetic variation arising during host infection. This strategy enabled the detection of convergent adaptation patterns at an unprecedented resolution. We also outline distinctive

eLife digest The bacterium *Staphylococcus aureus* lives harmlessly on our skin and noses. However, occasionally, it gets into our blood and internal organs, such as our bones and joints, where it causes severe, long-lasting infections that are difficult to treat.

Over time, *S. aureus* acquire characteristics that help them to adapt to different locations, such as transitioning from the nose to the blood, and avoid being killed by antibiotics. Previous studies have identified changes, or 'mutations', in genes that are likely to play an important role in this evolutionary process. One of these genes, called accessory gene regulator (or *agr* for short), has been shown to control the mechanisms *S. aureus* use to infect cells and disseminate in the body. However, it is unclear if there are changes in other genes that also help *S. aureus* adapt to life inside the human body.

To help resolve this mystery, Giulieri et al. collected 2,500 samples of *S. aureus* from almost 400 people. This included bacteria harmlessly living on the skin or in the nose, as well as strains that caused an infection. Gene sequencing revealed a small number of genes, referred to as 'adaptive genes', that often acquire mutations during infection. Of these, *agr* was the most commonly altered. However, mutations in less well-known genes were also identified: some of these genes are related to resistance to antibiotics, while others are involved in chemical processes that help the bacteria to process nutrients.

Most mutations were caused by random errors being introduced in to the bacteria's genetic code which stopped genes from working. However, in some cases, genes were turned off by small fragments of DNA moving around and inserting themselves into different parts of the genome.

This study highlights a group of genes that help *S. aureus* to thrive inside the body and cause severe and prolonged infections. If these results can be confirmed, it may help to guide which antibiotics are used to treat different infections. Furthermore, understanding which genes are important for infection could lead to new strategies for eliminating this dangerous bacterium.

---

signatures of adaptation during colonisation, upon transition from colonisation to infection and during invasive infection.

Results

The *S. aureus* within-host evolution analysis framework

We compiled a collection of 2251 *S. aureus* genomes from 267 independent episodes of colonisation and/or infection, reported in 24 genomic studies (*Young et al., 2017*; *Giulieri et al., 2018*; *Gao et al., 2015*; *Young et al., 2012*; *Wuthrich et al., 2019*; *Trouillet-Assant et al., 2016*; *Tan et al., 2019*; *Suligoy et al., 2018*; *Rouard et al., 2018*; *Rishishwar et al., 2016*; *Petrovic Fabijan et al., 2020*; *Miller et al., 2020*; *Loss et al., 2019*; *Liu et al., 2020*; *Langhanki et al., 2018*; *Kuroda et al., 2019*; *Ji et al., 2020*; *Howden et al., 2011*; *Harkins et al., 2018*; *Golubchik et al., 2013*; *Burd et al., 2014*; *Benoit et al., 2018*; *Azarian et al., 2019*; *Altman et al., 2018*; *Table 1*; *Table 1—source data 1*). We supplemented this dataset of publicly available sequences with unpublished sequences from 603 serial invasive isolates collected within the CAMERA-2 trial (*Tong et al., 2020*).

Using genetic distance and sequence type (ST) to define within-host lineages, we estimated that coinfection was present in 4/336 (1%) of invasive episodes and co-colonisation 11/167 (7%). We removed genetically unrelated strains within the same episode and included 2590 genomes (1397 invasive and 1193 colonising) from 396 episodes in our within-host evolution analysis (*Figure 1*, *Table 1*, *Supplementary files 1 and 2*). The most prevalent lineages in the collection were ST 30 (342 strains, 13%), ST 22 (277 strains, 11%), and ST 5 (271 strains, 11%); 1001 strains (39%) were *mecA* positive. The collection was representative of the global *S. aureus* diversity, with an even distribution of colonising and invasive strains across the major clades (*Figure 2A*). The most frequent infection syndrome was bacteraemia without focus (152 episodes, 38.4%), while nasal carriage (166 episodes, 42%) was the most prevalent colonisation condition (*Table 1*).

Our within-host evolution analysis strategy identified 4556 genetic variants (median 3 per episode, range 0–237) (*Supplementary file 3*). Importantly, by investigating both point mutations and structural variation, we were able to uncover 214 large deletions (≥500 bp), 160 new insertion sequences (IS)

Table 1. Microbiological and clinical characteristics of the colonisation and infection episodes included in the within-host evolution analysis.

Table 1—source data 1 provides a list of the within-host studies included in the analysis.

	Strains (n=2590)	Episodes (n=396)
Sequence type		
30	342 (13.2%)	43 (10.9%)
22	277 (10.7%)	44 (11.1%)
5	271 (10.5%)	42 (10.6%)
45	198 (7.6%)	38 (9.6%)
15	156 (6.0%)	4 (3.5%)
1	133 (5.1%)	14 (3.5%)
93	110 (4.2%)	29 (7.3%)
8	107 (4.1%)	18 (4.5%)
239	100 (3.9%)	29 (7.3%)
Other	896 (34.6%)	125 (31.6%)
mecA positive	1001 (38.6%)	207 (52.3%)
Infection syndrome		
Skin infection	204 (7.9%)	32 (8.1%)
Osteoarticular infection	77 (3.0%)	17 (4.3%)
Bacteraemia without focus	588 (22.7%)	152 (38.4%)
Bacteraemia with focus	331 (12.8%)	85 (21.5%)
Endocarditis	197 (7.6%)	44 (11.1%)
No invasive strains		66 (16.7%)
Colonisation syndrome		
Nasal carriage	974 (37.6%)	166 (42%)
Cystic fibrosis	57 (2.2%)	9 (2%)
Atopic dermatitis	162 (6.3%)	9 (2%)
No colonising strains		212 (54%)

The online version of this article includes the following source data for table 1:

Source data 1. List of within-host studies included in the analysis.

insertions, and 609 copy number variants, underscoring the role of large chromosome structural variation in within-host evolution. To increase the evolutionary convergence signal by aggregating mutations in functionally consistent categories, we annotated all genetic variants using multiple datasets, including coding sequences, regulatory intergenic regions, operons and gene ontologies (*Figure 1C*).

Distinctive evolutionary patterns define nose-adapted, early infection-adapted, and late infection-adapted strains

Based on the working hypothesis that *S. aureus* host adaptation patterns differ according to whether the strains are nose-colonising, collected at an early stage of infection (i.e. within the first 3 days) or at a late infection stage (i.e. associated with persistence beyond the first 3 days or recurrence), we assessed whether it was possible to define (i) general paradigms of genetic variation and (ii) specific convergence signatures. Thus, we classified within-host acquired variants into three groups according to their most likely location in the within-host phylogeny: (i) between colonising strains (colonising-colonising [type C>C]); (ii) between colonising and early infection adapted strains (colonising-invasive [type C>I]); and (iii) between invasive strains (invasive-invasive [type I>I]). Overall, the 396 infection episodes included in the analysis allowed us to independently assess 166 type C>C, 118 type C>I, and
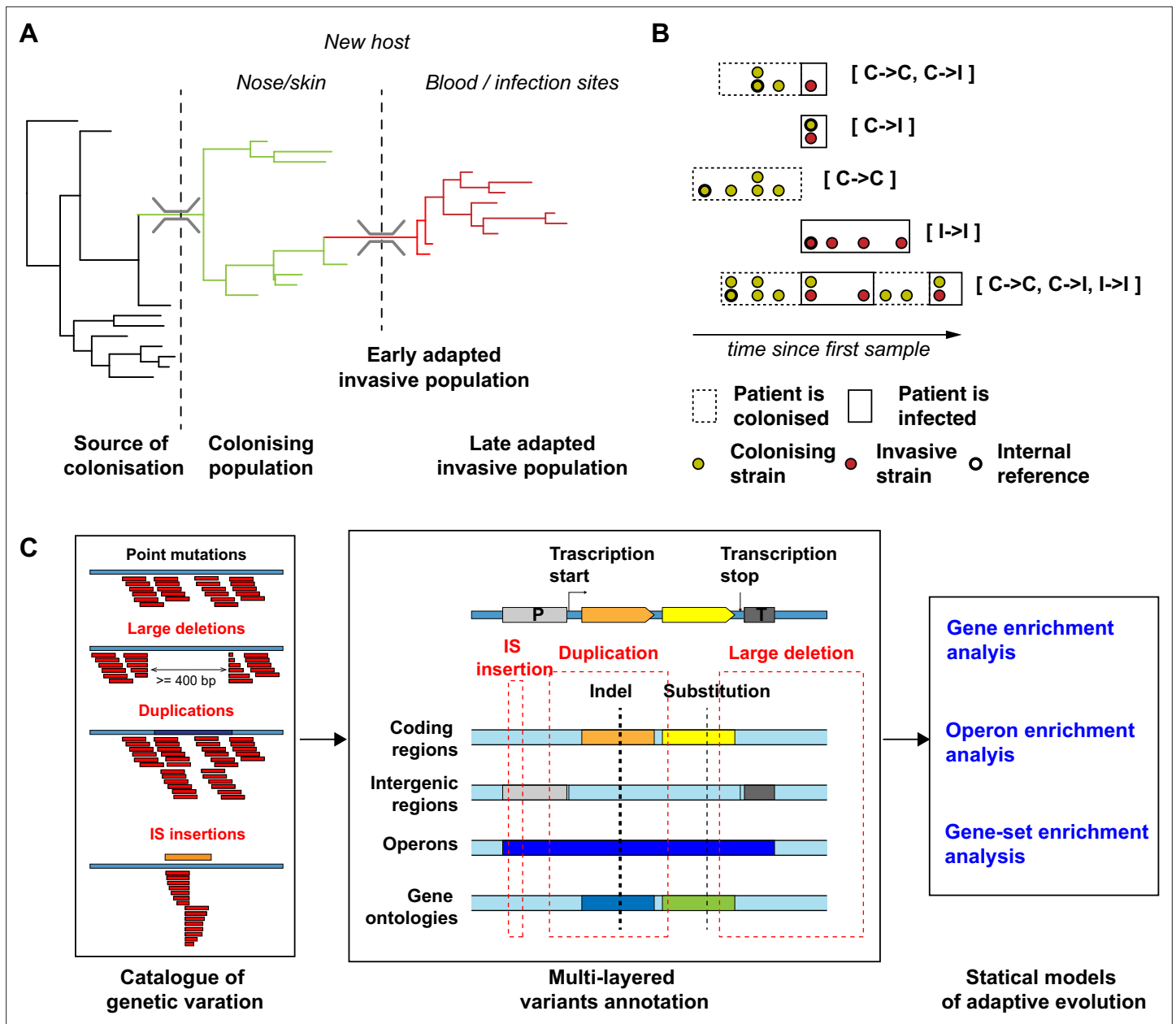
Figure 1. Overview of the *S. aureus* within-host evolution analysis framework. (**A**) Simulated phylogenetic tree illustrating within-host evolution of *S. aureus* colonisation and infection. This model assumes two genetic bottlenecks (dotted lines); upon transmission and upon transition from colonisation to invasive infection. (**B**) Sites and timing of within-host samples and number of genomes per sample define five prototypes of within-host evolution studies, each with colonising-colonising (C>C), colonising-invasive (C>I), or invasive-invasive (I>I) comparisons in different combinations: from top to the bottom: multiple colonising samples and one invasive samples; one colonising and one invasive sample; multiple colonising samples; multiple invasive samples; multiple colonising and invasive samples. (**C**) Approach to capture signals of adaptation across multiple independent episodes of colonisation/infection through detection of multiple genetic mechanisms of adaptation from short reads data and multi-layered functional annotation of the genetic variants using multiple databases including characterisation of intergenic regions (promoters), operon prediction, and gene ontology (GO). Statistical framework for the gene, operon, and gene set enrichment anlaysis (GSEA). Counts of independent mutations with likely impact on the protein sequence (non-synonymous substitutions, frameshifts, stop codon mutations, and insertion sequences [IS] insertions) were computed for each genes with a FPR3757 homologue. Gene counts (with the addition of intergenic mutations in promoter regions) were aggregated in operons and GOs. Gene and operon counts were used to fit Poisson regression models to infer mutation enrichment and significance of the enrichment. GOs counts and gene enrichment significance were used to run a gene-set-enrichment analysis. To illustrate the approach, the example of the gene *walR* is provided in italic.
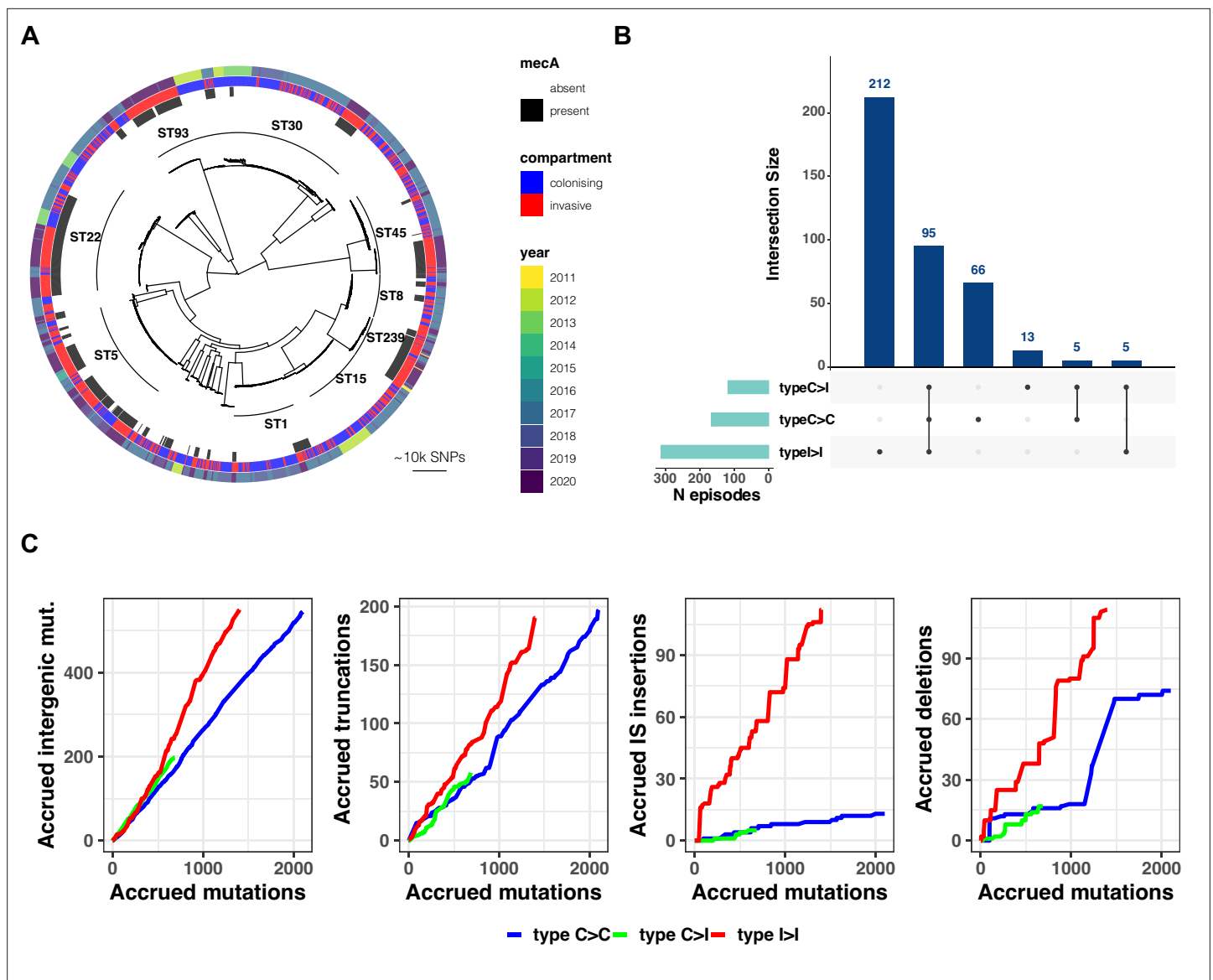
Figure 2. (**A**) Maximum-likelihood phylogenetic tree of 2590 *S. aureus* sequences included in the study. The tree is annotated (starting from the inner circle) with the most prevalent sequence types (ST), presence/absence of the *mecA* gene, compartment of isolation (colonising or invasive), and year of publication. (**B**) Summary of 396 independent episodes of *S. aureus* colonisation or infection categorised according to whether they allowed comparing colonising-colonising (C>C), colonising-invasive (C>I), or invasive-invasive (I>I) strains, or a combination of them. (**C**) Evidence of a distinctive pattern of adaptation in late infection-adapted strains (type I>I variants). For each type of comparison (type C>C, colonising-colonising; type C>I, colonising-invasive; type I>I, invasive-invasive), the cumulative curves display the accrued number of intergenic mutations, truncating mutations, insertion sequences (IS) insertions, and large deletions as a function of the total number of mutations. Genetic events were counted once per episode, regardless of the number of strains with the mutation. The sequence of mutations events in the cumulative curves is random.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Number of episode-specific variants in same-episode strains having the same sequence type (ST) as the internal reference vs. isolates with a different ST.

Figure supplement 2. Correlation between number of samples per episode and mean mutation counts.

Figure supplement 3. Within-host mutation rates within the colonising and invasive populations.

Figure supplement 4. Regression diagnostics to assess linear regressions sampling time after the internal reference and number of mutations.

Figure supplement 5. Distribution of new IS insertions by classification of the transposase and by major sequence types (ST).

312 type I>I within-host variants. In 95 cases, there were sufficient samples to assess all three types within the same episode (**Figure 2B**). Across colonisation/infection stages, sampling frequency did not seem to affect the number of variants identified with the exception of early adapted invasive strains (**Figure 2—figure supplement 2**).

We first sought to explore whether the rate of mutations over time differs between colonising and invasive populations. To estimate within-host mutation rates, we fitted linear regressions using data from a subset of 109 episodes that had at least two colonising or invasive isolates collected at two different timepoints (total 701 strains). The regressions suggested that mutation rates were higher in the invasive population; however, this analysis was limited by the heterogeneity of sampling strategies and by evidence of heteroskedasticity and non-linear distribution (**Figure 2—figure supplements 3 and 4**).

We have previously shown that invasive strains from persistent or relapsing infections exhibit a high proportion of protein-truncating mutations (**Giulieri et al., 2018**). A similar enrichment of protein-truncating variants was identified within invasive strains as compared to strains from asymptomatically colonised individuals (**Young et al., 2017**; **Young et al., 2012**). We reasoned that if this indicates genome degradation during infection, infecting strains might also be enriched for other loss of function (LOF) mutations caused by structural variants, such as movement of IS (**Hawkey et al., 2020**) and large deletions, leading to complete or partial gene loss (**Toft and Andersson, 2010**). In addition, we hypothesised that mutations and IS insertions in intergenic regions might contribute to altering gene expression or activity by interfering with the expression of key genes or operons (**McEvoy et al., 2013**).

Therefore, we calculated the prevalence of intergenic mutations, protein-truncating mutations, IS insertions, and large deletions among all variants and compared it between type C>C, type C>I, and type I>I variants. Strikingly, the distribution of mutations according to the predicted effect differed substantially in I>I pairs when compared to mutations identified between nose-colonising and invasive strains and within colonising strains (**Figure 2**). This can be expressed using the neutrality index (NI), which tests deviation from neutral evolution and is comparable to an odds ratio (**Stoletzki and Eyre-Walker, 2011**). Relative to type C>C variants, variants emerging within the infecting strains were enriched for intergenic mutations (NI 2.5; $p=1.8 \times 10^{-16}$) and protein-truncating mutations (NI 2.4; $p=4.8 \times 10^{-10}$) (**Table 2**). In contrast, no significant enrichment was observed among type C>I variants.

While large deletions were significantly more enriched in type I>I variants (NI 4.0, $p=1.1 \times 10^{-15}$), the strongest evidence for enrichment (NI 19.9, $p=1.6 \times 10^{-42}$) was found for IS insertions. We and others have previously shown that new insertions of IS256 may provide an efficient mechanism of genomic plasticity in invasive *S. aureus* strains (**Giulieri et al., 2018**; **Kuroda et al., 2019**; **McEvoy et al., 2013**). Here, we expand this observation in a larger dataset and show that this mechanism is not limited to IS256 (**Figure 2—figure supplement 5**). As shown in **Figure 2C**, two invasive strains exhibited a burst of >10 new IS insertions (IS3 and IS256, respectively). It has been shown that IS activation occurs under stress conditions, such as antibiotic exposure and oxidative stress (**Schreiber et al.,**

Table 2. Modified McDonald-Kreitman table displaying counts of variants (point mutations and structural variants) and the neutrality index for colonising-invasive (type C>I) and invasive-invasive (type I>I) variants (both compared to colonising-colonising [type C>C] variants).

Classification of variant	Number of variants *(Neutrality index)*		
	Type C>C	Type C>I	Type I>I
Synonymous	381	130	155
Non-synonymous	978	300 *(0.9)*	503 *(1.3)**
Intergenic	544	197 *(1.1)*	549 *(2.5)***
Truncating	197	58 *(0.9)*	190 *(2.4)***
Insertion sequences insertion	17	6 *(1.0)*	137 *(19.8)***
Large deletion	76	17 *(0.6)**	122 *(3.9)***

Values are counts of independent mutations. The neutrality index is shown in brackets in italic.
Significance testing Fisher's Exact Test: * p<0.05; ** p<0.005.

Figure 3. Top 20 genes with the most significant mutation enrichment across the entire dataset. (**A**) Significance of the enrichment for protein-altering mutations. The dashed line depicts the Bonferroni-corrected significance threshold, and red circles and blue circles represent genes with p values below and above the Bonferroni threshold, respectively. (B) Bar plots of independent mutations separated in three panels according to the type of variant (type C>C: colonising-colonising; type C>I: colonising-invasive; type I>I: invasive-invasive) and coloured according to the class of mutation. (C) Gene maps with type and positions of mutations.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Mapping of mutations in the 10 most significantly enriched mutated genes across the entire dataset.

Figure supplement 2. dN/dS values for non-synonymous mutations (**A**), indels (**B**), and non-sense mutations (stop codons) (**B**) for FPR3757 genes.

Figure supplement 3. Scatter plot representing in silico inferred functional impact of variants in the 20 most convergent loci.

Figure supplement 4. Most frequently deleted genes in large deletions.

Figure supplement 5. Most frequently enriched genes in copy number variations.

Figure supplement 6. Gene convergence analysis of all mutated genes (i.e. including both genes with FPR3757 homologue and no FPR3757 homologue).

Figure supplement 7. Gene convergence analysis after removing variants in strains included in *Young et al., 2017*, the largest collection of this analysis (1078 strains and 105 episodes).

2013), which is consistent with the selection environment encountered by invasive strains. However, these bursts occurred only in 2/1068 adapted invasive strains.

Overall, these data support a model, where late infection-adapted strains show an enrichment for variants that are predicted to exert a stronger functional impact, either by producing a truncated protein or by potentially interfering with intergenic regulatory regions, through point mutations or IS insertions. This strong genome degradation signature appears to be specific to type I>I variants and was absent in type C>I variants, suggesting that the bottleneck effect upon blood or tissue invasion does not explain it. To assess whether this general enrichment of non-silent evolution represented a signature of positive selection or derived from within-host gene obsolescence occurring during invasive infection, we further investigated signals of gene, operon, and pathway specific enrichment across independent episodes of infection.

Gene enrichment analysis identifies significant hotspots of adaptation

To identify signatures of adaptation, we first counted how many times each coding sequence was mutated independently across distinct colonisation/infection episodes (*Figure 3*, *Table 3*, *Table 4*, *Supplementary file 3*). We considered all protein-modifying mutations either predicted to cause a

Table 3. Genome-wide significant gene signatures of within-host evolution.
The genes shown reached genome-wide significance in the entire dataset or in either colonising-colonising (type C>C), colonising-invasive (type C>I), or invasive-invasive (type I>I) variants.

Gene	p value (whole dataset)	Description	N independent mutations			Significance
			Type C>C	Type C>I	Type I>I	
agrA*	7.04×10^{-28}	Accessory gene regulator protein A	5**	9**	8**	Part of the agr quorum sensing system, which is the master regulator of virulence factors expression in *S. aureus*. Recurrent mutations associated with invasive disease.
agrC**	2.84×10^{-10}	Accessory gene regulator protein C	4	2	6**	Histidine kinase, receptor for extracellular autoactivating peptide. Phosphorylates agrA.
stp1**	1.13×10^{-7}	Protein phosphatase 2 C domain-containing protein	3	2	3	Associated with vancomycin resistance.
mprF**	4.55×10^{-6}	Oxacillin resistance-related FmtC protein	2	0	9**	Main determinant of daptomycin resistance. Association with persistence and immune evasion.
rpoB	7.24×10^{-3}	DNA-directed RNA polymerase subunit beta	1	1	7**	Association with rifampicin resistance, but selection in the absence of rifampicin exposure can happen (R503H). Co-resistance to vancomycin, daptomycin, and oxacillin. Association with persistence.

*Significant enrichment (above the Bonferroni-corrected cut-off, see methods).

gain or LOF to the locus: non-synonymous substitutions, truncations, IS insertions, or deletions. To ensure consistency across the dataset, we restricted our analysis to 1736 (74%) genes with a homologue in reference strain FPR3757 (excluding plasmid genes and phage genes). Mutations were considered independent if they arose in distinct colonisation/infection episodes. To assess whether the convergent signals were a reliable indication of adaptation, we applied a gene enrichment analysis for protein-altering mutations which computes a length-corrected gene-level enrichment of protein-modifying mutations. The significance of the enrichment for each gene was estimated by comparing gene-specific Poisson models of mutation counts with the null hypothesis, which indicates neutrality and assumes a constant mutation rate across all genes (*Young et al., 2017*).

When applying a Bonferroni-corrected significance threshold (4.6×10^{-5}), mutations in *agrA* were highly significantly enriched across the entire dataset (45-fold enrichment, p=7.0×10^{-28}). Other significantly enriched genes were *agrC* (13-fold enrichment, p=2.8×10^{-10}), *stp1* (14-fold enrichment, p=1.1×10^{-7}), and *mprF* (sixfold enrichment, p=4.6×10^{-6}). The gene *sucA* reached near-significance (fivefold enrichment, p=6.8×10^{-5}). Mutations in genes most significantly targeted by convergent evolution were evenly distributed across the *S. aureus* phylogeny, indicating that these adaptative mechanisms were not specific to selected lineages (*Figure 3—figure supplement 1*). Using dN/dS analysis, we confirmed signatures of positive selection in the most significantly enriched genes, although only *agrA* reached statistical significance (*Figure 3—figure supplement 2* and *Supplementary file 4*).

We found that several genes with the most significant enrichment (*agrA*, *agrC*, *stp1*, and *sucA*) were recurrently mutated across all three within-host evolutionary scenarios, implying a global role in *S. aureus* adaptation during colonisation and invasion. This suggests partial adaptation of *S. aureus* strains upon invasion. It has been previously shown that adaptive mutations, particularly within the quorum sensing accessory gene regulator (*agr*), are enriched in invasive *S. aureus* strains, suggesting that adapted strains are more prone to be involved in invasive disease (*Young et al., 2017*; *Young et al., 2012*; *Altman et al., 2018*; *Smyth et al., 2012*). While the key role of *agr* was consistent with previous evidence from clinical and experimental studies, the high number of recurrent *sucA* mutations was surprising. This metabolic gene encodes for the α-ketoglutarate dehydrogenase of the tricarboxylic acid cycle, and recent work has revealed the functional basis of its potential role in adaptation. Its inactivation was found to lead to a persister phenotype (*Wang et al., 2018*), and *sucA* was a hotspot of metabolic adaptation to antibiotics in a recent in vitro evolution study (*Lopatkin et al., 2021*).

Adaptive mutations can cause both loss or gain of function of the gene affected; however, LOF is thought to be most frequent consequence of adaptation (*Behe, 2010*). We inferred LOF when variants lead to protein truncation (including intragenic IS insertions) or non-synonymous substitutions were expected to be deleterious based on the degree of divergence in conserved sites of the protein

Table 4. Gene signatures of within-host evolution with suggestive significant enrichment.

The genes shown reached the suggestive significance threshold in the entire dataset or in either type C>C, type C>I, or type I>I variants.

Gene	p value (whole dataset)	Description	N independent mutations			Significance
			Type C>C	Type C>I	Type I>I	
sucA*	6.82×10^{-5}	2-oxoglutarate dehydrogenase E1 component	6	2	2	Encodes a subunit of the α-ketoglutarate dehydrogenase of the tricarboxylic acid cycle.
saeR*	1.83×10^{-4}	DNA-binding response regulator SaeR	2	1	2	Regulator component of the saeRS two-component system. Virulence regulation.
accB	4.27×10^{-4}	Biotin carboxyl carrier protein of acetyl-CoA carboxylase	3*	1	0	Part of the fatty acid synthesis pathway of *S. aureus*.
SAUSA300_1856	6.41×10^{-4}	Hypothetical protein	4*	0	0	Intracellular cysteine peptidase. Putative chaperone in *S. aureus*.
xpaC	1.38×10^{-3}	Hypothetical protein	4*	0	0	Predicted 5-bromo-4-chloroindolyl phosphate hydrolysis protein, no data on *S. aureus*.
rpsJ	1.58×10^{-3}	30S ribosomal protein S10	3*	0	0	Mutations at residues 53–60 are associated with tigecycline resistance, at no apparent fitness cost.
SAUSA300_2399	1.68×10^{-3}	ABC transporter ATP-binding protein	4*	0	0	Downregulated in the presence of fusidic acid
walR	2.10×10^{-3}	DNA-binding response regulator	1	0	3*	Part of walKR two-component response regulator. Associated with vancomycin resistance.
yjbH	3.55×10^{-3}	Dsba-family protein	1	0	3*	Negative regulator of spx (directs its ClpXP-dependent degradation). Association with antibiotic resistance, virulence regulation, and oxidative stress resistance.
purR	3.86×10^{-3}	Pur operon repressor	0	1	3*	*purR* mutants: increased biofilm formation and virulence in animal model; higher capacity to invade epithelial cells.
era	5.34×10^{-3}	GTP-binding protein Era	0	1	3*	Involved in ribosome assembly and stringent response.
pbp2	7.75×10^{-3}	Penicillin-binding protein 2	6*	0	0	Role in methicillin resistance (PBP2a synergism). Increased expression after oxacillin exposure.
fakA	9.90×10^{-3}	Hypothetical protein	5*	0	0	Fatty acid kinase. Deletion mutant displayed increased virulence in a murine model of skin infection.
sgtB	2.65×10^{-2}	Glycosyltransferase	0	0	3*	sgtB mutations in adaptive laboratory evolution experiments upon vancomycin exposure.

*suggestive significant enrichment (above the suggestive significance cut-off, adjusted for false-discovery, see methods).

Choi et al., 2012. The majority (63%) of aggregated variants in adaptive loci were predicted to lead to LOF (*Figure 3—figure supplement 3*), either because of accumulation of truncating mutations (*agrA and agrC*) or exclusively because of deleterious substitutions (*walR and vraG*). Because only LOF is inferred from the sequence, gain of function variants might be miss-classified as neutral (or even deleterious). We considered this hypothesis for genes with low frequency of truncations and expected adaptive advantage of gain of function mutations based on the literature. For example, *mprF* mutations associated with vancomycin or daptomycin resistance have been shown to be associated with increased enzymatic activity of the protein leading to decreased negative charge of the membrane (*Ernst and Peschel, 2019*). Among within-host acquired *mprF* mutations in our dataset, we found that four of them were previously described and associated with antibiotic resistance through a gain of function mechanism (*Ernst and Peschel, 2019*). A distinctive sign of these variants was convergence at position or mutation level, which has been described as a potential hallmark of gain of function mutations (*Gerasimavicius et al., 2021*). Further supporting the evidence for gain of function, *mprF* was duplicated in one of the episodes, as reported previously (*Gao et al., 2015*), while no other convergent gene was affected by within-host copy number variants. We also assessed *accB* because no variant in this genes was classified as deleterious; however, *accB* mutations have been previously described as LOF in strains that are auxotroph for fatty acids (*Morvan et al., 2016*). Thus, based on in silico prediction and previous data, we hypothesise that variants in convergent loci were mostly expected to be LOF with the exception of *mprF*.
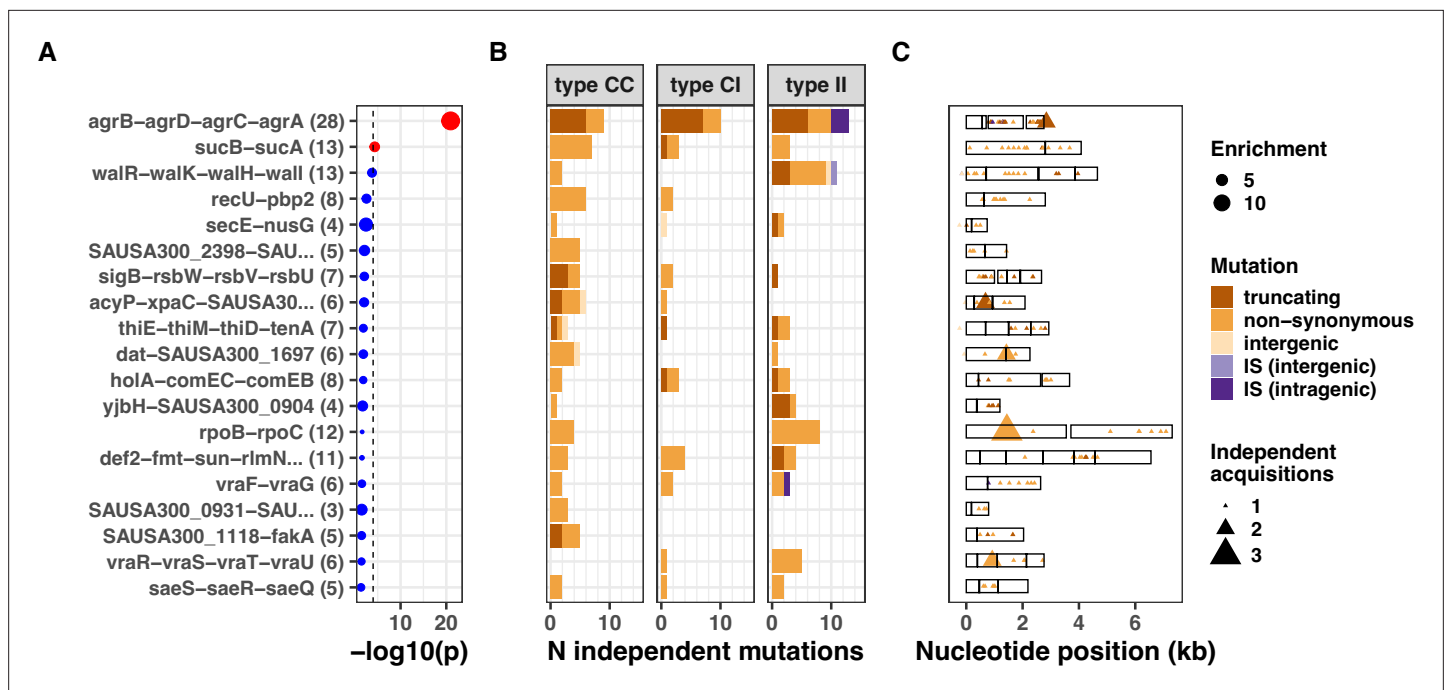
Figure 4. Top 20 operons with the most significant mutation enrichment across all dataset. (**A**) Significance of the enrichment for protein-altering mutations. The dashed line depicts the Bonferroni-corrected significance threshold, and red circles and blue circles represent operons with p values below and above the Bonferroni threshold, respectively. (**B**) Bar plots of independent mutations separated in three panels according to the type of variant (type C>C: colonising-colonising; type C>I: colonising-invasive; type I>I: invasive-invasive) and coloured according to the class of mutation. Mutations were considered independent if they occurred in separate episodes of either colonisation or invasive infection. (**C**) Operon maps with positions of the mutations (relative to the start of the first gene of the operon). Operons are labelled with the names of the genes included, and longer labels were shorted for clarity (see **Supplementary file 5** for details).

To confirm that our gene enrichment analysis (focused on point mutations and IS insertions and limited to genes with FPR3757 homologues) captured the large part of adaptation, we analysed variation due to large deletions and copy number variation, which were not included in the gene enrichment analysis. We observed multiple independent deletions and amplifications mainly in phage genes (**Figure 3—figure supplement 4** and **Figure 3—figure supplement 5**). We also repeated the gene enrichment analysis with all mutated genes (with and without FPR3757 homologues) and found very similar results, with only two hypothetical proteins with no FPR3757 homologue among the genes with most significant enrichment (**Figure 3—figure supplement 6**).

Combining multiple mechanisms of adaptation and multi-layered annotation increases the signal of convergent evolution

To increase our ability to capture signatures of adaptation from convergent evolution, we extended our analysis beyond coding sequences, to integrate the genetic variation signals issued from intergenic mutations and IS insertions in intergenic regions. This multi-layered annotation of mutated regions was shown to increase the amount of information gained from in vitro adaptive evolution experiments (**Phaneuf et al., 2020**). Such methodology allows for an advanced classification of intergenic mutations based on regulatory sequences including promoters and transcription units based on data acquired from RNAseq experiments (**Mäder et al., 2016**; **Prados et al., 2016**).

Using this approach, we were able to assign 150/1237 (11%) of all intergenic mutations and IS insertions to a predicted regulatory region. We found that the *agr*, *sucAB*, and *walKR* operons had the strongest convergent evolution signal with 28, 13, and 13 independent mutations (**Supplementary file 5**). Mutations within these loci were significantly (*agr*, 12-fold enrichment, p=1.1e-21; *sucAB*, fourfold enrichment, p=4.5e-5) or near-significantly enriched (*walKR*, threefold enrichment, p=1.7e-4) (**Figure 4**). Interestingly, promoter mutations represented 2/13 (15%) of the *walKR* operon mutations, indicating that potentially impactful intergenic variants may be missed when considering only
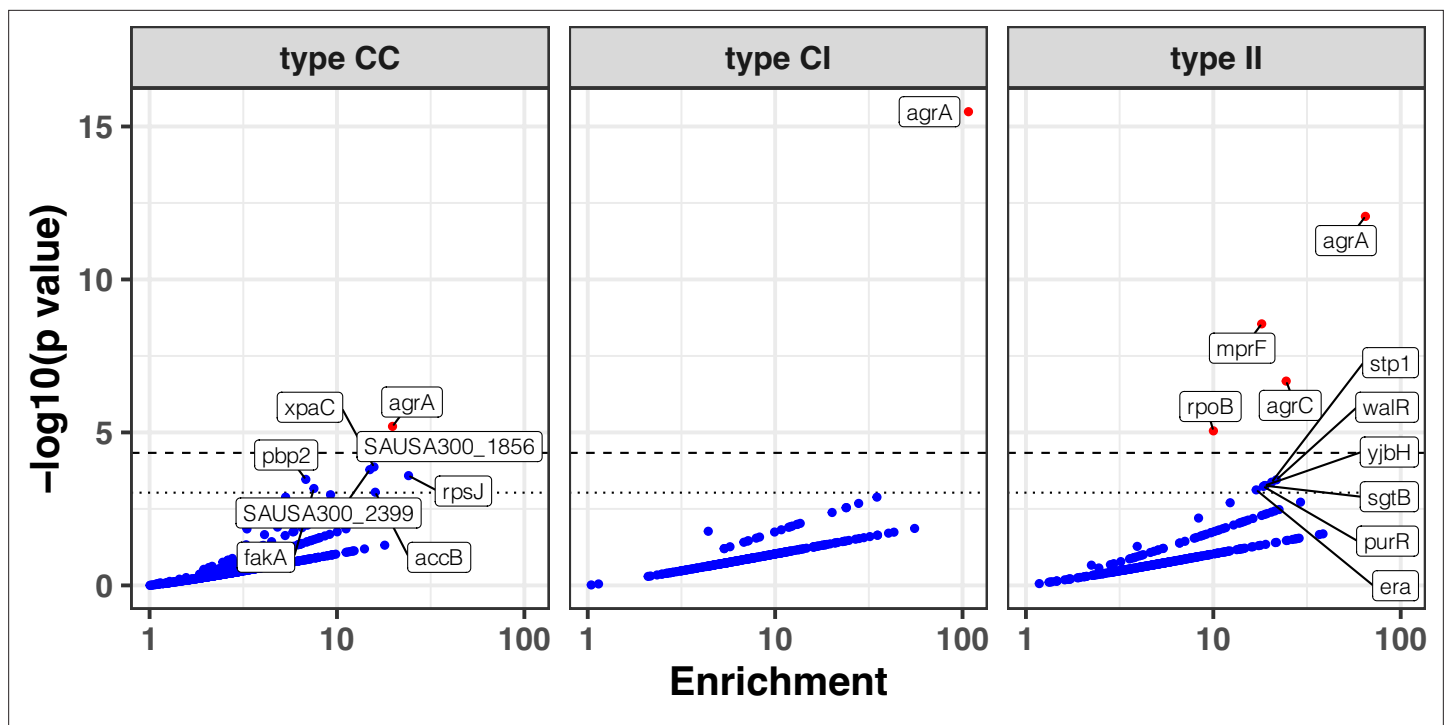
Figure 5. Modified volcano plot displaying enrichment (x-axis) and significance of enrichment (y-axis) within colonising-colonising (type C>C), colonising-invasive (type C>I), and invasive-invasive (type I>I) variants. The horizontal dashed line depicts the Bonferroni-corrected significance threshold and dotted line shows the suggestive significance threshold. Labels indicate genes with significance of enrichment below the suggestive threshold. Genes are coloured in red if the p value is below the Bonferroni-corrected threshold and in blue otherwise.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Modified volcano plot displaying enrichment (x-axis) and significance of enrichment (y-axis) for FPR3757 operons across the entire dataset, colonising-colonising (type CC), colonising-invasive (type CI), and invasive-invasive (type II) variants.

Figure supplement 2. Gene set enrichment analysis (GSEA) for protein-modifying mutations in colonising-colonising (type CC), colonising-invasive (type CI), and invasive-invasive (type II) variants.

coding regions. Furthermore, new IS insertions were found within type I>I variants: three insertions into *agrC* (predicted to inactivate the gene, as shown previously in staphylococci [***Both et al., 2021***; ***Suligoy et al., 2020***]) and an insertion 159 bp upstream of *walR*, in a region encompassing its cognate promoter. Together with the strong enrichment for IS insertions within type I>I variants, the location of these insertions in recurrently mutated operons suggests that IS insertions contribute to the adaptive evolution of *S. aureus* during invasive infection.

Adaptation within the invasive population is distinctive and strongly driven by antibiotics

The excess of non-silent evolution (and potentially function-altering) within invasive strains suggested that strong, specific selection pressure occurs within the invasive populations (type I>I variants). We therefore assessed genes that appeared to be specifically mutated or inactivated during infection. We performed our gene- and operon-enrichment analysis for each type of within-host variants separately (i.e. within the colonising population, between colonising and invasive strains, and within the invasive population) (***Figure 5***). We found that *agrA* mutations were highly enriched in any group of variants, and particularly prevalent between colonising and invasive strains (type C>I variants), consistent with a previous study that is included in this analysis (***Young et al., 2017***). Among type I>I variants (between invasive strains), a significant enrichment was observed in *mprF* (18-fold enrichment, $p=2.8 \times 10^{-9}$), *agrC* (24-fold enrichment, $p=2.1 \times 10^{-7}$), and *rpoB* (10-fold enrichment, $p=8.8 \times 10^{-6}$). Other genes that were strongly enriched in type I>I variants (below the Bonferroni-corrected threshold, but above the suggestive significance threshold, ***Figure 5***) included *walR* (22-fold enrichment, $p=3.5 \times 10^{-4}$), *stp1*

(20-fold enrichment, p=4.2 × 10⁻⁴), *yjbH* (19-fold enrichment, p=5.4 × 10⁻⁴), *sgtB* (19-fold enrichment, p=5.5 × 10⁻⁴), and *purR* (18-fold enrichment, p=5.8 × 10⁻⁴).

The enrichment for mutations in *mprF, rpoB, stp1, sgtB,* and in the *walKR/yycH* operon (11-fold enrichment, p=9 × 10⁻⁹, see *Figure 5—figure supplement 1* for the operon enrichment analysis) highlights the role of antibiotic pressure in shaping adaptation within the invasive population, since these loci are hotspots of adaptation to key anti-staphylococcal antibiotics that are often used in invasive infections (rifampicin, daptomycin, and vancomycin). For example, the essential two-component regulator *walKR/yycFG* (and its associated genes *walH/yycH*) have been shown to have a key role in vancomycin resistance in one of the within-host evolution studies included in this analysis (*Howden et al., 2011*), while mutations in both *stp1* and *sgtB* have been observed in vancomycin-adapted strains (*Machado et al., 2021*).

Notably, the most significant gene signatures in invasive strains might have been selected in response to other selective pressures, including the host immune response during infection. For example, *rpoB* mutations have been associated with pleiotropic effects, including co-resistance to vancomycin, daptomycin, and oxacillin and immune evasion, suggesting a potential role in adaptation beyond the response to the selective pressure from rifampicin (*Guérillot et al., 2018*). This hypothesis is supported by the presence of mutations (such as the *rpoB* R503H substitution and N405 inframe deletion) outside the rifampicin-resistance determining region.

Pleiotropic phenotypes are also likely to underlie the enrichment of *yjbH* with invasive strains, which was mutated four times (of which three were truncations), yet only one mutation was found in colonising strains or early infection-adapted strains. This gene has a cysteine-rich domain that is homologous to *dsbA* in *Escherichia coli*. One of its roles in *S. aureus* is to facilitate the ClpXP-dependent degradation of the transcriptional regulator Spx (*Austin et al., 2019*). Inactivation of *yjbH* has been associated with oxacillin (*Göhring et al., 2011*) and vancomyin (*Renzoni et al., 2011*) resistance, impaired growth (*Engman et al., 2012*), and reduced virulence in animal models (*Paudel et al., 2021*), indicating that *yjbH* mutations may influence both host-pathogen interaction and antibiotic resistance. Finally, *purR*, a purine biosynthesis repressor, has been recently characterised beyond its metabolic function: interestingly, it was shown to be a virulence regulator (*Sause et al., 2019*), where *purR* mutants displayed higher bacterial counts following mice infection, increased biofilm formation (*Goncheva et al., 2019*), and higher capacity to invade epithelial cells (*Goncheva et al., 2020*).

We performed a gene set enrichment analysis (GSEA), using gene ontology and antibiotic resistance gene annotations (*Feldgarden et al., 2019*). The GSEA, stratified by variant type, showed significant enrichment only in type I>I variants, further underscoring the higher level of adaptation in this group (*Figure 5—figure supplement 2* and *Supplementary file 6*) and confirmed the broad functional implications of the most enriched genes and operons with the invasive populations, since among the ontologies that were significantly enriched within the invasive population, we found the categories 'DNA binding' (normalised enrichment score [NES]=1.6, false discovery rate [FDR]-adjusted p=9 × 10⁻⁴), 'pathogenesis' (NES = 1.7, adjusted p=4 × 10⁻³), and 'antibiotic response' (NES = 1.8, adjusted p=7 × 10⁻³).

Taken together, these findings point to six key genetic loci that appear to have an important role in *S. aureus* adaptation during invasive infections. These loci are associated with either antibiotic resistance (*mprF, rpoB, stp1, sgtB,* and *walKR*), pathogenesis (*agrAC* and *purR*), or both (*yjbH*).

A mutation's co-occurrence network defines loci under within-host co-evolutionary pressure

Epistasis, defined as the interaction of multiple mutations on a given phenotype (*Levin-Reisman et al., 2019*), plays a role in adaptive evolution in bacteria, particularly in antibiotic resistance (*Skwark et al., 2017*; *Wadsworth et al., 2018*; *Yokoyama et al., 2018*). Whether epistatic interactions could promote *S. aureus* adaptation during infection remains unknown. Identifying these interactions would enable identification of combinations of mutations underlying bacterial adaption during infection and refine the prediction of infection outcomes. Here, we assessed co-occurrence of mutations and mutated genes across independent episodes of colonisation/infection. While co-occurrence may simply result from co-selection (e.g. simultaneous exposure to two different antibiotics), it may also indicate putative epistatic interactions that could be explored in terms of potential impact on adaptive phenotypes (*Phillips, 2008*).
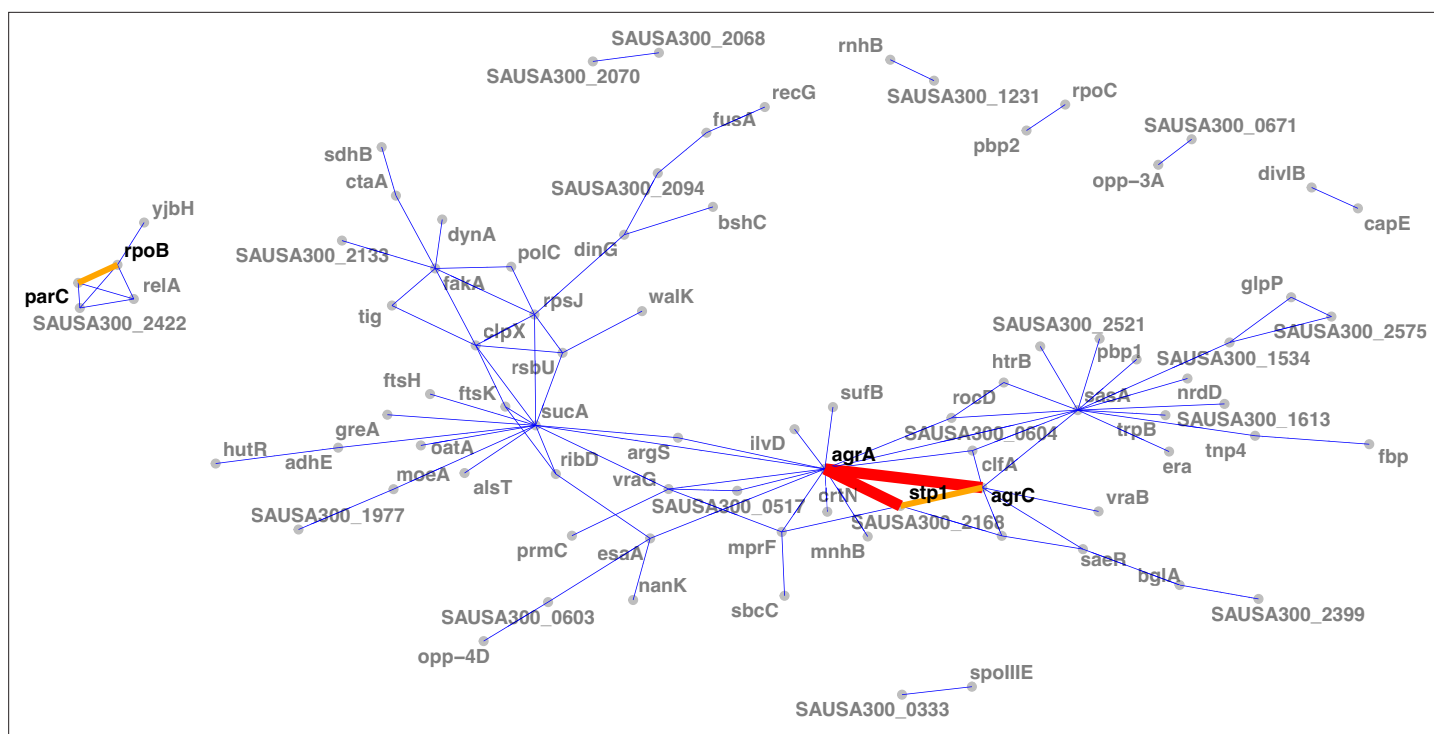
Figure 6. Network of mutations co-occurrence. The width and colour of the edges represent the strength of the co-occurrence of mutated genes on the same strain (thin and blue, two independent co-occurrences; thick and orange, three independent co-occurrences).

First, we explored co-occurrence of mutations and found only one case where the same mutations co-occurred in more than one independent episode. The two mutations were an inframe deletion within hypothetical protein SAUSA300_2068 and a A60D substitution of the gene *ywlC*. These genes are closely located in FPR3757. While the co-occurrence could be explained by recombination, recombination is expected to be rare amongst within-host *S. aureus* populations in general (*Golubchik et al., 2013*) and even rarer within invasive strains. YwlC is a threonylcarbamoyl-AMP synthase in *E. coli*, thus it is possible that SAUSA300_2068 is also a ribosomal protein. Ribosomal proteins can display regulatory activity (*Aseev and Boni, 2011*) and could plausibly be targets of adaptation to both antibiotics and the host/intracellular survival. This specific case of convergent co-occurrence of mutations was detected within type I>I variants.

When assessing interactions at gene level (i.e. co-occurrence of the same altered protein sequences across independent episodes), we found the strongest interaction between the *agrA* and *agrC* genes (*Figure 6*). While this is consistent with the high convergence of mutations in the *agr* locus, this suggests that strains acquire multiple mutations within the locus, possibly further impacting *agr* activity. Interestingly, no convergent co-occurrence signature compatible with possible epistasis was observed within the *walKR* locus, the other operon with a high number of independent mutations; which could be due to the essentiality of *walKR* in *S. aureus* (*Monk et al., 2019*). Collectively, *agr* locus mutations interacted with 17 other mutated genes, the strongest interaction being with *stp1*. Since *stp1* (a serine/threonine phosphatase) has been previously associated with virulence regulation (*Cameron et al., 2012*), this interaction potentially indicates another mechanism by which adapted strains fine-tune the gene expression profile that is already altered by *agr* mutations.

Another moderately strong interaction was observed between *rpoB* and *parC*, which were co-mutated in three independent episodes. Given the association of *parC* mutations with fluoroquinolone resistance (*Trong et al., 2005*), this interaction is likely to be an example of co-selection due to co-exposure to fluoroquinolones and rifampicin.

Clinical correlates of adaptive signatures within colonising and invasive populations

Genetic signatures of bacterial adaptation have been associated with infection extent, for example, enabling the prediction of extraintestinal infection with *S. enterica* (*Wheeler et al., 2018*). We have previously shown that adaptive mutations are enriched in invasive infections; however, it is unclear whether bacterial adaptation is more likely to be associated with distinctive clinical syndromes. To identify clinical correlates of adaptive signatures, we classified colonisation and infection episodes based on the sites of collection and on clinical data obtained from the publications (*Table 1* and *Figure 7—figure supplement 1*). We then used the Jaccard index and network analysis to compute node centrality as a global measure of adaptation for each independent episode. The Jaccard index can be used as a simple marker of the proportion of shared mutated genes between pairs of colonisation or infection episodes (*Bailey et al., 2017*). Node centrality allows to similtuaneously take into account the strength of similarity between independent episodes (Jaccard index) and the number of pairs with shared mutated genes (number of connections). Hence, a node centrality of 0, indicates that the episode does not share any mutated genes with other episodes and appears as isolates node on the adaptation network (*Figure 7—figure supplement 2*). Here, we limited the analysis to the 20 most significantly enriched genes with each type of variant.

Our network analysis showed that adaptation was present in only a minority of episodes within each type of variant (*Figure 7—figure supplement 2*). With a definition of adaptation based on a centrality value of more than 0, we found that the proportion of adaptive episodes was 43, 20, and 22% with type C>C, C>I, and I>I variants, respectively. In addition, certain clinical syndromes were more strongly associated with adaptation. Within the colonising population (type C>C variants), almost 80% of cystic fibrosis episodes were adaptive, as opposed to one third of episodes of skin colonisation in atopic dermatitis (*Figure 7AB*). This is consistent with within-host evolution studies showing strong convergent evolution signals among bacterial populations colonising individuals with cystic fibrosis, not only in case of *S. aureus* colonisation (*Long et al., 2020*) but also *P. aeruginosa* (*Marvig et al., 2015*) and *Mycobacterium abscessus* (*Bryant et al., 2021*); however, one study found adaptive evolution signals in atopic dermatitis (*Key et al., 2021*). We also observed that adaptation among infection episodes correlated with infection extent. Episodes of infective endocarditis episodes displayed higher adaptation metrics (46% with centrality >0) than bacteraemia with additional infection foci (28%) and bacteraemia without focus (17%) (*Figure 7D–E*).

To explore the syndrome-specificity of adaptation signatures, we mapped mutations in the most significantly enriched adaptive genes to clinical syndromes of colonisation and infection (*Figure 7* panels C and F). As expected, syndromes with high prevalence of adaptation had higher numbers of episodes with adaptive mutations; however, some genes appeared to be preferentially mutated. For example, *rpsJ*, *stp1,* and SAUSA300_1230 were over-represented in cystic fibrosis, while no clear pattern of mutations was discernible for nasal carriage episodes. Within infection syndromes, *mprF* and *purR* mutations were more prevalent in endocarditis, and *yjbH* mutations were only found in severe infections (bacteraemias with additional foci and endocarditis). Some genes appeared to be distinctive for low adaptation groups (atopic dermatitis and skin infections); however, the low number of adaptative mutations prevented an accurate assessment of these profiles.

Discussion

Within-host evolution of bacterial pathogens such as *S. aureus* is thought to be governed by a combination of positive selection for variants that confers an advantage within the host and random fixation of mutations (genetic drift) (*Klemm et al., 2016*; *Didelot et al., 2016*). Sudden changes in the effective population size (bottlenecks) can cause genetic drift, for example, when a single or few bacterial cells invade the bloodstream or when a secondary infection foci is established in tissues and organs. Consistent with this view, animal studies have shown that after infecting the blood with a polyclonal population, bacteraemic infection is established stochastically by a single clone (*McVicker et al., 2014*; *Guerillot et al., 2018*), although estimating the bottleneck size at invasion from clinical sequences has been more challenging (*Abel et al., 2015*). On the other hand, several lines of evidence support the role of positive selection and adaptive evolution during *S. aureus* infection. First, adaptive phenotypic features appear to be acquired during infection. The most obvious adaptative
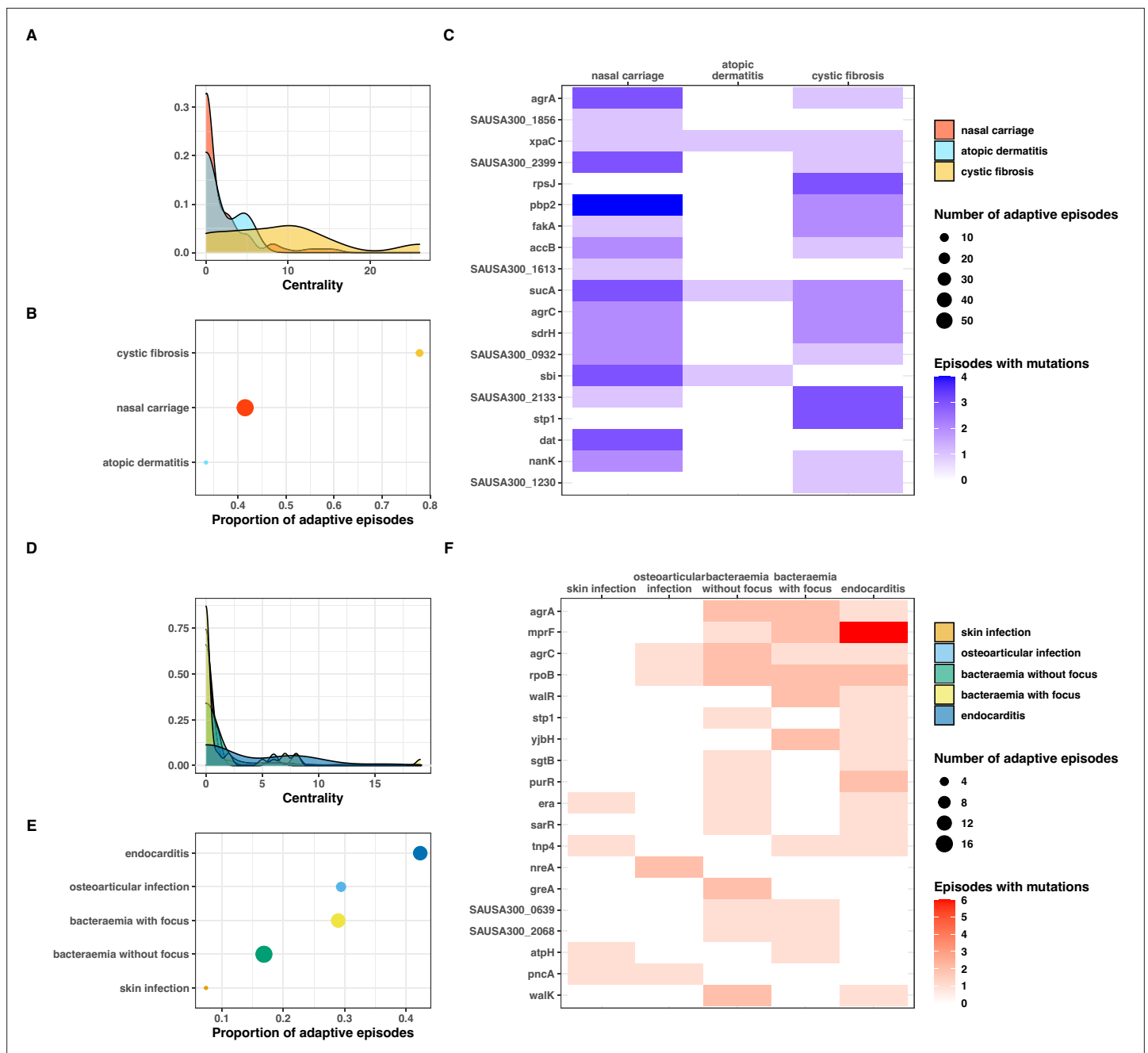
Figure 7. Clinical correlates of adaptive signatures within colonising (colonising-colonising [type C>C,] panels A–C) and invasive (invasive-invasive [type I>I], panels D–F) bacterial populations. Adaptation was inferred by computing the Jaccard index of shared mutated genes between independent episodes, followed by network analysis of infection episodes pairs. The node centrality measure was used as an indicator of adaptation. To avoid overinflation of mutated genes, the calculation was limited to the 20 most significantly enriched genes within each group of mutations. (**A, D**) Density of centrality values across colonisation (panel A) and infection categories (panel D). (**B, E**) Number and proportion of adaptive episodes. An adaptive episode was defined by a centrality >0. (**C, F**) Distribution of mutations in the 20 most significantly enriched genes across categories of colonisation (panel C) and infection (panel F).

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. Clinical manifestations and infection sites of invasive episodes, grouped by the infection syndromes classification used for the adaptation analysis.

Figure supplement 2. Network of colonisation/infection episodes for colonising-colonising (type CC) (panel A), colonising-invasive (type CI) (panel B), and invasive-invasive (type II) variants (panel C).

phenotype is secondary resistance to antistaphylococcal antibiotics such as rifampicin, vancomycin (*Howden et al., 2006*), daptomycin (*Peleg et al., 2012*), and oxacillin (*Giulieri et al., 2020*). Crucially, these resistance phenotypes can be associated with pleiotropic, pathoadaptative phenotypes such as small colony variant and immune evasion (*Guérillot et al., 2018*; *Jiang et al., 2019*; *Guérillot et al., 2019*). Furthermore, phenotypic adaptation (e.g. loss of toxicity) has been observed upon transition from colonisation to infection (*Laabei et al., 2015*), supporting the concept that invasive infection is linked to pathoadapted strains. At the molecular level, an excess of protein-truncating mutations in invasive strains (*Giulieri et al., 2018*) and in late colonising strains leading to infection (*Young et al., 2012*) have been noted. While this observation alone could be explained by relaxed constraint resulting from reduced population size (*Didelot et al., 2016*), it has been suggested that loss of gene function might be a common adaptation mechanism of within-host evolution (*Gatt and Margalit, 2021*), as supported by evidence of gene- or pathway-specific enrichment of mutations across independent infection episodes (*Young et al., 2017*).

Despite support for adaptive evolution from previous studies, it has been difficult to identify specific molecular signatures of adaptation during infection, due to the limited power of previous within-host studies of bacteraemia and other serious *S. aureus* infections that were often limited to a restricted number of episodes. To increase our ability to identify signatures of adaptation and find significantly enriched loci, we analysed multiple sources of genetic variation (point mutations, large deletions, IS insertions, and copy number variants) in a large collection of independent episodes of *S. aureus* colonisation and infection from 25 studies. We predicted that the main advantage of our approach would be to increase the ability of detecting convergence of genetic variants arising during invasive infections as opposed to those detected during the colonisation and upon transition from colonisation to infection. To test this hypothesis, we classified within-host variants based on their likely position in the within-host phylogeny (*Figure 1D*).

Bacterial adaptation is promoted by genomic plasticity; however, within-host evolution is characterised by low genetic variation (*Giulieri et al., 2018*). Based on our previous genomic studies of *S. aureus* bacteraemia, we reasoned that chromosome structural variants may provide an additional mechanism to increase genetic variation during infection. Here, we found that new insertions of IS are strongly enriched during invasive infection. However, despite this 20-fold enrichment, IS insertions remained a rare source of variation even within invasive strains and appeared to have a selective contribution to adaptation (i.e. limited to specific loci such as *agrC*). Similarly, large deletions and copy number variants appeared to play a less prominent role in adaptation, although we did not include them in our enrichment analysis.

Together with the enrichment for LOF mutations, which is another feature of evolution within the invasive population in our analysis, these IS movements suggest that a pattern of reductive evolution (genome degradation through loss of genes or accumulation of LOF mutations) emerges during within-host evolution of invasive *S. aureus*. This genome degradation might be related to less effective purifying selection (loss of deleterious alleles) in the invasive population due to a decrease in effective population size and a shorter evolutionary timescale (*Didelot et al., 2016*). However, our data indicate that these changes converge to specific genes, operons, and pathways, suggesting an adaptative benefit. Reductive evolution has been described in several 'commensal-to-pathogen' settings (*Toft and Andersson, 2010*). Although niche adaptation through reductive evolution has been described in extracellular pathogens (*Stinear et al., 2007*), a smaller (reduced) genome is a hallmark of obligate intracellular endosymbiotic bacteria (*Batut et al., 2014*). Since it appears that invasive *S. aureus* is able to reside intracellularly (promoting dissemination through mobile phagocytes during *S. aureus* sepsis [*Thwaites and Gant, 2011*; *Surewaard et al., 2016*], and immune evasion), it is plausible that this pattern of reductive evolution reflects an adaptation of invasive *S. aureus* to an invasive and intracellular lifestyle, as it has been shown for other facultative intracellular pathogens such as non-typhoidal *Salmonella* (*Klemm et al., 2016*). However, it is possible that these signatures of reductive evolution might be only temporary, as genome degradation might be present only in a minority of strains or be reversible; moreover, LOF mutations are expected to be more likely than gain of function mutations.

Beside reductive evolution, another distinctive feature of within-host evolution during invasive infection was intergenic mutations (both point mutations and IS insertions). In a within-host evolution study of *S. pneumoniae* colonisation, it was shown that intergenic sites were under convergent evolution (*Chaguza et al., 2020*). Mutations in promoter sequences of some core genes can play an

important role in antibiotic resistance as it was repeatedly shown for *pbp4* and resistance to beta-lactam antibiotics (*Basuino et al., 2018*). The role of intergenic mutations in within-host evolution was shown in a study of *P. aeruginosa* infection, where convergent evolution targeted several intergenic regulatory regions including upstream of antibiotic resistance genes (*Khademi et al., 2019*).

Previous work on within-host evolution by our group (included in this analysis) has established that *agrA* mutations are significantly enriched upon transition from colonisation to infection (*Young et al., 2017*). In addition, we have shown through genomics and targeted mutagenesis that mutations in key genes such as *walKR* (*Howden et al., 2011*) and *rpoB* (*Gao et al., 2013*) play a key pathoadaptive role in selected cases of persistent *S. aureus* infections. In this study, we increased our ability to discover potential targets of adaptation by analysing several mechanisms of genetic variation and applying several layers of annotation. As compared to previous work on *S. aureus*, this approach provides a higher-resolution picture of within-host evolution and adaptation. Importantly, this analysis remains robust after removing more than 1000 sequences from the largest within-host study included (*Figure 3—figure supplement 7*). We increase here the generalisability of our findings. We expand the list of genes targeted by convergent evolution and show that there are distinctive adaptation pathways in colonising and invasive populations. We confirm that the dominantly mutated loci belonged to the *agr* locus, in particular *agrA* and *agrC*. This finding is consistent with a large body of literature that predated the genomic era (*Novick and Geisinger, 2008*) that supports the role of the *agr* locus as the master regulator of gene expression in *S. aureus*. *Agr*-mediated adaptation was so important that we found a highly significant enrichment of *agr* mutations arros all type variants, including within colonising strains (type C>C variants). Shopsin et al. showed that ~10% of healthy *S. aureus* carriers held an agr-defective strain and that prior hospitalisation was significantly associated with agr-defective status, suggesting prior adaptive pressures (*Shopsin et al., 2008*).

Consistent with the distinctive general patterns of evolution displayed during invasive infection, some genes and loci were specifically mutated within invasive strains. Some of these genes were emerging targets of *S. aureus* pathogenesis in vivo, such as *purR* and *yjbH* that were not singled out in previous within-host evolution investigations. Others were known determinants of antibiotic resistance, including *mprF*, *rpoB,* and the *walKR* operon. This underscores the crucial role of antibiotic exposure in shaping adaptive evolution during invasive infection. A recent study of within-host evolution during cystic fibrosis found that resistance genes were hotspots of convergent evolution in this population, which is frequently treated with antibiotics and shows features of phenotypic adaptation (*Long et al., 2020*).

While most mutations in adaptive loci were substitutions within the coding sequence, about 40% of *walKR* operon mutations were located outside of the coding regions of *walR* and *walk*, emphasising the need to study intergenic mutations and mutations throughout an operon to capture adaptive signatures. This observation highlights the importance of expanding the analysis of intergenic mutations for the detection of adaptive mutations, in particular those linked to antibiotic resistance.

If within-host evolution represents adaptative evolution, it is possible that adaptation involved an accumulation of mutations and possibly epistatic mechanisms. Our data show that some mutations are specific for invasive strains; these mutations may reflect late adaptation, occurring after evolution during colonisation in the nose and upon transition from colonisation to infection, and thus occurring after adaptive mutations were acquired during earlier stages. This evolutionary pattern of stepwise adaptation (or adaptive continuum) encompassing the entire within-host evolutionary arch has been well described for cancer (*Abbosh et al., 2017*) and has been also investigated in a study of transition from colonisation to infection (*Young et al., 2012*). One way to capture this is mutation co-occurrence analysis. Here, we show mutation co-occurrence within the *agr* operon but also co-occurrence of the same mutations in two uncharacterised proteins in *S. aureus*. Our network of mutation co-occurrence linked to *agrAC* mutations might suggest a potential pathway of stepwise adaptation following initial mutations acquisition in the *agr* locus, an hypothesis that has been explored in one of studies included in our analysis (*Altman et al., 2018*).

While combining multiple studies allowed us to increase statistical power in order to detect genome-wide convergent signals of adaptation, this approach has some limitations. The quality of the publicly available sequences and metadata can be heterogeneous, despite performing quality control assessment, for example, due to different read coverage across studies. Lack of consistent metadata might have impacted the clinical categorisations used here, including the distinction between

colonising and invasive strains (however, 83% of the strains could be classified unambiguously based on the site of collection). In addition, detection of structural variants from short reads is not as accurate as from long reads; for example, chromosomal inversions can be missed if their inversion site span is larger than the insert site of the paired-end reads (*Guérillot et al., 2019*). Furthermore, the majority of the episodes included had low sampling density. More contemporary strategies leverage a multiplicity of samples and deep sequencing strategies to capture within-host diversity, allowing to detect minority variants that could be relevant for adaptation and to obtain a more accurate understanding of evolutionary dynamics within the host, including estimating bottlenecks (*Hall et al., 2019*) and tracking the fall and rise of within-host lineages (*Bryant et al., 2021*). Finally, while several adaptive loci identified here have been previously assessed experimentally, the functional impact of adaptive variants in less characterised genes and intergenic mutations warrants further exploration using targeted mutagenesis (*Monk et al., 2015*).

Ultimately, the goal of detecting adaptive signals is to identify new mechanisms of pathogenesis or resistance to therapeutic targets and to inform prediction of clinical outcomes. So far, studies have failed to show consistent associations between specific clinical outcomes and genetic features of infecting (or colonising) *S. aureus* strains. This might be related to the dominance of host/environmental factors, but it could also be linked to the limited power of studies performed so far. By contrast, within-host evolution studies are an elegant approach to identify signatures of adaptation that might be candidate markers of important clinical outcomes, such as infection risk in case of colonisation or treatment failure in case of infection. Here, we show that adaptation signatures are at least partially specific to colonisation, infection, and upon transition from colonisation to infection and that adaptive changes are more frequent in distinctive infection episodes (complicated bacteraemia and endocarditis). These findings suggest that adaptive signatures might be indicative of important clinical outcomes. In the future, precision medicine in infectious diseases could follow the lead of cancer genomics, where within-host evolution studies have tracked the evolution of cancer clones and enabled the detection of high risk mutations early.

Materials and methods

Literature search

We conducted a search of articles indexed in PubMed before the 11 August 2020 using the keyword 'aureus' in combination with either 'genomics' or 'whole genome sequencing' and with either 'within-host evolution', '*in vivo* evolution', 'adaptation', or 'bacteraemia'. The records retrieved through this search were combined with additional citations identified through other sources. After removing duplicates, this resulted in 815 citations that were screened based on following inclusion criteria: (i) whole-genome sequencing of human *S. aureus* isolates; (ii) >1 *S. aureus* isolates sequenced per individual; (iii) sequences (reads or assemblies) publicly available; and (iv) minimum sequences metadata available (either with the manuscript or linked to the sequences): patient ID, date of collection (or collection interval in reference to a baseline isolate), and source of the sample. After excluding studies not satisfying the inclusion criteria (730 based on the title, 46 based on the abstract, and 15 after reviewing the full text), we kept 24 within-host evolution studies.

Extraction of sequence metadata

For each of the included studies, the following variables were extracted either from reads, metadata (when available) or from the publication/supplementary data: identifier linking the sequences to a patient or an episode of infection, date of collection (when available) or collection interval in reference to a baseline isolate, and site of collection of the isolate. Isolates were broadly categorised in 'colonising' and 'invasive' based on the site of collection, when the information was unambiguous (e.g. 'nose' for 'colonising' or 'blood' for 'invasive'). When the information on the body site was not sufficient (e.g. 'skin' or 'lung'), the categorisation was based on further details provided in the publication. When available, phenotypic metadata and antibiotic treatments were also extracted from the publication. We used clinical and site data to classify colonisation episodes in 'nasal carriage', 'atopic dermatitis', and 'cystic fibrosis' and infection episodes in 'skin infection' (skin infection site surgical site infection without other foci), 'osteoarticular infection' (bone/join infection without other foci), 'bacteraemia without focus' (bloodstream infection, no other foci, and expect for vascular catheter or skin),

'bacteraemia with focus' (bloodstream infection with other focus involving the lung, nervous system, bone and joints, or internal organs), and 'endocarditis' (based on diagnosis reported in the publication or in the clinical metadata).

Sequence processing

Sequences (reads and assemblies) and metadata were downloaded from the European Nucleotide Archive and the National Center for Biotechnology Information (NCBI), respectively using the BioProject accession or the genome accession. Quality assessment of the reads was performed by calculating mean read depth and the fraction of *S. aureus* reads using Kraken 2, v2.0.9-beta (*Wood et al., 2019*) and by extracting metrics from reads assemblies constructed using Shovill, v1.1.0 (https://github.com/tseemann/shovill, *Seemann, 2022c*) and annotated using Prokka, v1.14.6 *Seemann, 2014* . ST was inferred from the assembly using Mlst, v2.19.0 (https://github.com/tseemann/mlst, *Seemann, 2022b*), and resistance genes were detected using Abricate, v1.0.1 (https://github.com/tseemann/abricate, *Seemann, 2022a*). Reads were discarded if the mean coverage depth was below 35, the majority of reads were not *S. aureus,* or the size of the assembly was below 2.6 megabases. Assemblies downloaded from the NCBI repository were discarded if the genome size was below 2.6 megabases.

Sequences from the CAMERA2 trial

We collected *S. aureus* strains from bacteraemia episodes included in the CAMERA2 trial (Combination Antibiotics for Methicillin Resistant *S. aureus*), where at least two strains per episode were available. The CAMERA2 trial was performed between 2015 and 2018 in Australia, New Zealand, Singapore, and Israel and randomised participants with methicillin-resistant *S. aureus* bacteraemia to either monotherapy with vancomycin or daptomycin or combination therapy with vancomycin or daptomycin plus an antistaphylococcal beta-lactam (flucloxacillin, cloxacillin, or cefazolin) (*Tong et al., 2020*). Strains were isolated from –80C glycerol onto horse-blood agar. Species were confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Bacterial whole-genome sequencing was performed from single colonies on the Illumina NextSeq platform. Reads were checked for quality, assembled, and annotated as described above.

Global phylogeny

To generate a global alignment of all sequences, reads and shredded assemblies were mapped to reference genome USA300 FPR3757 (assembly accession: GCF_000013465.1) (using Snippy, v4.6.0) (https://github.com/tseemann/snippy; *Seemann, 2022d*). The core genome alignment was obtained using Snippy; sites with >10% gaps were removed using Goalign (*Lemoine and Gascuel, 2021*) and constant sites were removed using SNP-sites (*Page et al., 2016*), for a final length of 186,825 bp. A maximum-likelihood phylogenetic tree of 2590 sequences (those kept in the analysis after excluding genetically unrelated strains, see below) was inferred using IQ-TREE, v2.0.3 *Minh et al., 2020*.

Variant calling

We have previously shown that the accuracy of variant calling in within-host evolution analyses is improved when mapping reads to an internal draft assembly as opposed to a closely related closed genome (*Giulieri et al., 2018*). Here, we applied the same approach, where we selected the internal reference among the sequences from the same patient or episode. When available, the oldest colonising strain was selected. When only sequences from invasive strains were available, the oldest strain (baseline strain) was selected. When multiple sequences were available per sample (e.g. multiple colonies sequenced per plate) or at the same date, the reference was randomly selected among them. Snippy with default parameters (minimum reads coverage 10, minimum read mapping quality 60, and minimum base quality 13) was used for variant calling. To further improve the accuracy of the calls, we masked variants called from reference reads and those at positions where reference reads had a coverage below 10 (using the BEDTools suite [*Quinlan and Hall, 2010*]).

Filtering of genetically unrelated sequences

The threshold for removing genetically unrelated sequences was set empirically at 100 episode-specific variants based on the upper Tukey's fence of the distribution of the number of variants in same-episode isolates belonging to the same ST (*Figure 2—figure supplement 1*).

Estimation of within-host mutation rates

To estimate within-host mutation rates within colonising and invasive populations, a linear regression was fitted to model the relationship between sampling time (in years after the first isolate) and number of mutations relative to the internal reference. Only episodes with at least two strains collected at least one day apart were included in this analysis. The mutation rate μ was computed as follows $\mu = \beta/g$, where β is the regression parameter and g is the mean genome size of the internal references (2.79 Mb). Regression diagnostics were performed using the R package performance (*Lüdecke et al., 2021*).

Detection of chromosome structural variants

Using BWA-MEM (*Li, 2013*), reads and shredded contigs were aligned to the closest available complete genome (either internal to the dataset or selected from the NCBI repository based on the mash distance). To detect large deletions (≥ 500 bp), reads coverage was computed using BEDTools, as described in *Giulieri et al., 2018*. To detect new IS insertions, split reads were extracted, filtered, and annotated as described in *Giulieri et al., 2018*. We used the R package CNOGpro (*Brynildsrud et al., 2015*) to detect 1000 bp windows with an estimated copy number above one as compared to the internal reference. The package calculates the reads coverage per sliding windows of the chromosome, performs a G+C bias normalisation, and infers copy number state using a Hidden Markow Model. We ran the package with default parameters, with the exception of the length of the sliding window that was set at 1000 bp. For each class of structural variant and within each episode, we used BEDTools to mask regions where the variant was already present in internal reference.

Prediction of functional impact of variants

Functional impact of variants was extracted from the Snippy output, which uses SnpEff to infer the functional effect of the detected mutations (*Cingolani et al., 2012*). SnpEff categories for coding regions were aggregated in 'truncating' (frameshift, stop codons, and start codons), 'non-synonymous substitutions', and 'synonymous substitutions'. Non-synonymous substitutions were further investigated using PROVEAN, v1.1.5 (58) using the non-redundant protein database provides on the PROVEAN repository (ftp://ftp.jcvi.org/pub/data/provean/nr_Aug_2011/). Substitutions were classified as 'deleterious' if the PROVEAN score was –2.5 or less and 'neutral' otherwise.

Internal variant annotation

To ensure a consistent annotation of mutated genes across independent episodes, we clustered amino-acid sequences using CD-HIT, v4.8.1 with an identity threshold of 0.9. The BEDTools suite was used to annotate mutated intergenic regions with upstream and downstream coding regions and the distance separating the mutation from the start or the end of the gene. For the operon analysis, intergenic mutations were classified according to their location within a presumed promoter based on blasting the sequence of unique promoters (as determined in [*Prados et al., 2016*]) on the draft assembly of the internal reference. Phage genes were annotated using blastp and the PhageWeb database (http://computationalbiology.ufpa.br/phageweb/).

Variant annotation using reference strain FPR3757

To compare mutated genes across separated episodes, we used blastp to identify homologues of each CD-HIT cluster of mutated genes in USA300 FPR3757. Genes in FPR3757 were further annotated using the database provided in the AureoWiki repository (*Fuchs et al., 2018*), and operon annotations of FPR3757 were retrieved from Microbes Online (*Dehal et al., 2010*). In addition, we used the text mining tool PaperBLAST to search for publications containing data on homologues of uncharacterised FPR3757 proteins (*Price et al., 2017*). Only protein-altering variants in genes with FPR3757 homologues (excluding plasmid genes and phage genes) were kept for the analysis of convergence at gene and operon level and the gene enrichment analysis.

Classification of variants

Mutational and structural variants were classified in to type C>C (within colonising strains), type C>I (between colonising and invasive strains), and type I>I (within invasive strains) as follows: all variants arising in colonising strains were classified as type C>C, while variants among invasive strains were classified as type C>I if they were found in a baseline invasive strain (defined as the oldest invasive strain; when multiple sequences were available at same time, the baseline invasive strain was selected randomly), and as type I>I if they were found between invasive strains but not on the baseline invasive strain. This approach is based on the assumption that co-infection or superinfection is rare, as we have shown previously for bacteraemia (*Giulieri et al., 2018*).

Calculation of the Neutrality Index (NI)

A modified McDonald-Kreitman table was compiled a described in *Stoletzki and Eyre-Walker, 2011*, where a ratio was calculated between non-synonymous, protein-truncating, IS insertions, intergenic and deletion variants, and synonymous variants. The NI was obtained by dividing the ratio calculated above for type C>I and type I>I by the ratio for type C>C variants that were used as reference group. Significance was tested by Fisher's Exact test.

dN/dS analysis

We used the R package dNdScv (*Martincorena et al., 2017*) to obtain dN/dS ratios for non-synonymous mutations, indels, and missense mutations (stop codons) for all FPR3757 genes, based on variants called when mapping all reads on FPR3757 and after subtracting variants from the internal reference reads and variants in positive where internal reference reads had a low coverage. Since this analysis could be hampered by potential false-positive variants resulting from mapping reads on a single reference (*Giulieri et al., 2018*), we also used our curated list of within-host mutations obtained from episode-specific variant calling to calculate crude dN/dS ratios by dividing the number of protein modifying mutations by the number non-synonymous mutations and computed p values by Fisher exact test as in *Long et al., 2020*.

Gene and operon enrichment analysis

We calculated the enrichment of protein-altering mutations across all coding regions of FPR3757 (excluding plasmid genes and phage genes) using the approach described in *Young et al., 2017*. The variant enrichment per gene i was calculated as follows: $(N_i/L_i)/(\Sigma_n/\Sigma_l)$, where N_i is the number of variants per gene i, L_i is the length of gene i, Σ_n is the total number of variants, and Σ_l is the total length of the genes. We used Poisson regression to model the number of variants per gene j under the null hypothesis (no enrichment), as defined by the equation $\lambda_0 L_j$, where λ_0 is the expected number of variants in any gene and L_j is the gene length. Under the alternative hypothesis (enrichment of variant in gene i), the estimated number of variants is $\lambda_i L_i$ for gene i, and $\lambda_1 L_j$ for any other gene j. The model parameters λ_0, λ_1, and λ_l were obtained using maximum likelihood and tested for significance using the likelihood ratio test. The genome-wide significance cut-off was calculated using the Bonferroni correction (0.05 divided by the number of unique genes or operons) and the suggestive significance cut-off (1 divided by the number of unique genes or operons), as implemented for bacterial genome-wide associated studies in *Lees et al., 2017a*.

Gene set enrichment analysis

We used the PANNZER platform (*Törönen et al., 2018*), to retrieve a gene ontology annotation of FPR3757 based on the GO terms. We modified the 'antibiotic response' category by adding a curated list of antibiotic resistance genes downloaded from the NCBI Anti-Microbial Resistance (AMR) gene reference database (*Feldgarden et al., 2019*). The GSEA was performed as implemented in the R package clusterProfile (*Yu et al., 2012*). Genes with a FPR3757 homologue were ranked according to the significance of the enrichment of protein-modifying mutations (gene enrichment analysis, see above), and the GSEA was carried out with a minimum gene set size of 10 and using the FDR method for adjustment for multiple testing.

Mutation co-occurrence analysis

To detect co-occurrence of mutations and mutated genes across independent episodes, we constructed a co-occurrence matrix using the R package co-occur (*Griffith et al., 2016*). A

co-occurrence of mutations or mutated genes in at least two independent episodes was interpreted as convergent and as a sign of potential epistatic interaction. The network of co-occurrence of mutated genes was visualised using the R package ggraph (https://cran.r-project.org/web/packages/ggraph/index.html).

Network analysis of adaptation signatures

The pairwise calculation of the Jaccard index between set of mutated genes was performed in R. The calculations were performed both with the entire set of mutated FPR3757 genes and with the 20 most significantly enriched genes in each group of variants. A network of shared mutated genes between independent episodes was constructed using ggraph, where edges represented episode connections based on the Jaccard index. We used the R package tidygraph to extract the node centrality (function centrality_degree) as a summary measure of the degree of adaptation of the episodes. The network graph and analysis were performed for each group of variants separately.

Acknowledgements

Additional information

Funding

Author contributions

Stefano G Giulieri, Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review and editing; Romain Guérillot, Conceptualization, Formal analysis, Methodology, Software, Supervision, Writing – review and editing; Sebastian Duchene, Resources, Software, Writing – review and editing; Abderrahman Hachani, Investigation, Writing – review and editing; Diane Daniel, Data curation, Investigation, Project administration; Torsten Seemann, Methodology, Software, Supervision; Joshua S Davis, Investigation, Resources, Writing – review and editing; Steven YC Tong, Conceptualization, Investigation, Resources; Bernadette C Young, Methodology, Writing – review and editing; Daniel J Wilson, Methodology, Software, Writing – review and editing; Timothy P Stinear, Conceptualization, Methodology, Supervision, Writing – review and editing; Benjamin P Howden, Conceptualization, Funding acquisition, Supervision, Writing – review and editing

Author ORCIDs
Abderrahman Hachani (iD) http://orcid.org/0000-0001-8032-2154
Steven YC Tong (iD) http://orcid.org/0000-0002-1368-8356
Bernadette C Young (iD) http://orcid.org/0000-0001-6071-6770
Daniel J Wilson (iD) http://orcid.org/0000-0002-0940-3311
Timothy P Stinear (iD) http://orcid.org/0000-0003-0150-123X
Benjamin P Howden (iD) http://orcid.org/0000-0003-0237-1473

Ethics

Ethics approval was obtained at each partecipating site to the CAMERA2 trial and written informed onsent was obtained from each participant or surrogate decision maker.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.77195.sa1
Author response https://doi.org/10.7554/eLife.77195.sa2

Additional files

Supplementary files

• Supplementary file 1. List of colonisation/infection episodes included with publication data (first author, year, PubMed id), number of strains, sites of collection, clinical characterstics, classification of colonisation, and infection episodes.

• Supplementary file 2. List of strains included with site and date of collection, sequence type, presence of the *mecA* gene, information on whether the strain was designed as internal reference or baseline index strain, mash distance to the internal reference, number of variants called (as compared to the internal reference), and sequencing metrics.

• Supplementary file 3. List of variants identified annotated with gene, gene sequence, FPR3757 homologue, and FPR3757 operon. Point mutations, insertion sequences insertions, large deletions, and copy number variants are presented separately.

• Supplementary file 4. Gene enrichment analysis for all mutated genes with a FPR3757 homologue with number of mutations, gene length, mutation enrichment, and p value based on a Poisson regression to model the number of variants per gene. Results are presented separately for the complete dataset and for colonising-colonising (type C>C), colonising-invasive (type C>I), and invasive-invasive (type I>I) variants.

• Supplementary file 5. Operon enrichment analysis for all FPR3757 operons (i.e. mutated genes that could be assigned to a FPR3757 operon) with number of mutations, operon length, mutation enrichment, and p value based on a Poisson regression to model the number of variants per operon. Results are presented separately for the complete dataset and for colonising-colonising (type C>C), colonising-invasive (type C>I), and invasive-invasive (type I>I) variants.

• Supplementary file 6. Gene set enrichment analysis for mutations in genes aggregated in gene ontologies (GO) categories with enrichment score, normalsied enrichment score (NES), and unadjusted and false-discovery rate (FDR) adjusted p value. Results are presented separately for the complete dataset and for colonising-colonising (type C>C), colonising-invasive (type C>I), and invasive-invasive (type I>I) variants.

• MDAR checklist

Data availability

All data generated or analysed during this study are included in the manuscript and supporting file 1-6. The code to call, filter and annotated within-host variants and to perform the enrichment analysis is available on github at https://github.com/stefanogg/staph_adaptation_paper, (copy archived at swh:1:rev:6ec5132855405e1c759fedadb4c70e295c1c1974).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Giulieri SG, Daniel D, Seemann T, Davis JS, Tong SYC, Stinear TP, Howden BP	2022	Within-host evolution analysis of *S. aureus* bacteraemias included in the CAMERA-2 trial	https://github.com/ stefanogg/staph_ adaptation_paper	European Nucleotide Archive, PRJEB50796

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Gao W, Monk IR, Tobias NJ, Gladman SL, Seemann T, Stinear TP, Howden BP	2015	Large tandem chromosome duplications facilitate niche adaptation during persistent infection with drug-resistant *Staphylococcus aureus*	https://www.ncbi.nlm. nih.gov/bioproject/ PRJEB9193	NCBI BioProject, PRJEB9193
Howden BP	2011	Evolution of Multidrug Resistance during *Staphylococcus aureus* Infection Involves Mutation of the Essential Two Component Regulator WalKR	https://www.ncbi.nlm. nih.gov/sra/?term= SRA027352	NCBI Sequence Read Archive, SRA027352
Howden BP	2011	Evolution of Multidrug Resistance during *Staphylococcus aureus* Infection Involves Mutation of the Essential Two Component Regulator WalKR	https://www.ncbi.nlm. nih.gov/bioproject/ PRJNA29567	NCBI BioProject, PRJNA29567
Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, Robins-Browne R, Davies JK, Seemann T, Stinear TP	2011	Evolution of Multidrug Resistance during *Staphylococcus aureus* Infection Involves Mutation of the Essential Two Component Regulator WalKR	https://www.ncbi.nlm. nih.gov/bioproject/ PRJNA29569	NCBI BioProject, PRJNA29569
Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CLC, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, Wilson DJ	2012	Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease	https://www.ncbi.nlm. nih.gov/bioproject/ PRJEB2862	NCBI BioProject, PRJEB2862
Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, Fung R, Godwin H, Knox K, Votintseva A, Everitt RG, Street T, Cule M, CL Ip, Didelot X, Peto TE, Harding RM, Wilson DJ, Crook DW, Bowden R	2013	Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage	https://www.ncbi.nlm. nih.gov/bioproject/ PRJEB2881	NCBI BioProject, PRJEB2881

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Burd EM, Alam MT, Passalacqua KD, Kalokhe AS, Eaton ME, Satola SW, Kraft CS, Read TD	2014	Development of oxacillin resistance in a patient with recurrent *Staphylococcus aureus* bacteremia	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248678	NCBI BioProject, PRJNA248678
Rishishwar L, Kraft CS, Jordan IK	2016	Population Genomics of Reduced Vancomycin Susceptibility in *Staphylococcus aureus*	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA259799	NCBI BioProject, PRJNA259799
Trouillet-Assant S, Lelièvre L, Martins-Simões P, Gonzaga L, Tasse J, Valour F, Rasigade JP, Vandenesch F, Muniz Guedes RL, Ribeiro de Vasconcelos AT, Caillon J, Lustig S, Ferry T, Jacqueline C, Loss de Morais G, Laurent F	2016	Adaptive processes of *Staphylococcus aureus* isolates during the progression from acute to chronic bone and joint infections in patients	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA298748	NCBI BioProject, PRJNA298748
Rouard C, Garnier F, Leraut J, Lepainteur M, Rahajamananav L, Languepin J, Ploy MC, Bourgeois-Nicolaos N, Doucet-Populaire F	2018	Emergence and Within-Host Genetic Evolution of Methicillin-Resistant *Staphylococcus aureus* Resistant to Linezolid in a Cystic Fibrosis Patient	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA434495	NCBI BioProject, PRJNA434495
Langhanki L, Berger P, Treffon J, Catania F, Kahl BC, Mellmann A	2018	In vivo competition and horizontal gene transfer among distinct *Staphylococcus aureus* lineages as major drivers for adaptational changes during long-term persistence in humans	https://www.ncbi.nlm.nih.gov/bioproject/PRJEB22600	NCBI BioProject, PRJEB22600
Altman DR	2018	Genome Plasticity of agr-Defective *Staphylococcus aureus* during Clinical Infection	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA393749	NCBI BioProject, PRJNA393749
Giulieri SG	2018	Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent *Staphylococcus aureus* bacteraemia	https://www.ncbi.nlm.nih.gov/bioproject/PRJEB27932	NCBI BioProject, PRJEB27932
Benoit JB, Frank DN, Bessesen MT	2018	Genomic evolution of *Staphylococcus aureus* isolates colonizing the nares and progressing to bacteremia	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA414752	NCBI BioProject, PRJNA414752
Suligoy CM, Lattar SM, Noto Llana M, González CD, Alvarez LP, Robinson DA, Gómez MI, Buzzola FR, Sordelli DO	2018	Mutation of Agr Is Associated with the Adaptation of *Staphylococcus aureus* to the Host during Chronic Osteomyelitis	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA414566	NCBI BioProject, PRJNA414566

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Harkins CP, Pettigrew KA, Oravcová K, Gardner J, Hearn RMR, Rice D, Mather AE, Parkhill J, Brown SJ, Proby CM, Holden MTG	2018	The Microevolution and Epidemiology of *Staphylococcus aureus* Colonization during Atopic Eczema Disease Flare	https://www.ncbi.nlm.nih.gov/bioproject/PRJEB20148	NCBI BioProject, PRJEB20148
Tan X	2019	Chronic *Staphylococcus aureus* Lung Infection Correlates With Proteogenomic and Metabolic Adaptations Leading to an Increased Intracellular Persistence	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA446073	NCBI BioProject, PRJNA446073
Loss G, Simões PM, Valour F, Cortês MF, Gonzaga L, Bergot M, Trouillet-Assant S, Josse J, Diot A, Ricci E, Vasconcelos AT, Laurent F	2019	*Staphylococcus aureus* Small Colony Variants (SCVs): News From a Chronic Prosthetic Joint Infection	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA497214	NCBI BioProject, PRJNA497214
Kuroda M, Sekizuka T, Matsui H, Ohsuga J, Ohshima T, Hanaki H	2019	IS256-Mediated Overexpression of the WalKR Two-Component System Regulon Contributes to Reduced Vancomycin Susceptibility in a *Staphylococcus aureus* Clinical Isolate	https://www.ncbi.nlm.nih.gov/bioproject/PRJDB8056	NCBI BioProject, PRJDB8056
Azarian T, Ridgway JP, Yin Z, David MZ	2019	Long-Term Intrahost Evolution of Methicillin Resistant *Staphylococcus aureus* Among Cystic Fibrosis Patients With Respiratory Carriage	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA527261	NCBI BioProject, PRJNA527261
Wüthrich D, Cuénod A, Hinic V, Morgenstern M, Khanna N, Egli A, Kuehl R	2019	Genomic characterization of inpatient evolution of MRSA resistant to daptomycin, vancomycin and ceftaroline	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA488707	NCBI BioProject, PRJNA488707
Ji S, Jiang S, Wei X, Sun L, Wang H, Zhao F, Chen Y, Yu Y	2020	In-Host Evolution of Daptomycin Resistance and Heteroresistance in Methicillin-Resistant *Staphylococcus aureus* Strains From Three Endocarditis Patients	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA577181	NCBI BioProject, PRJNA577181
Miller CR, Dey S, Smolenski PD, Kulkarni PS, Monk JM, Szubin R, Sakoulas G, Berti AD	2020	Distinct Subpopulations of Intravalvular Methicillin-Resistant *Staphylococcus aureus* with Variable Susceptibility to Daptomycin in Tricuspid Valve Endocarditis	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA544229	NCBI BioProject, PRJNA544229
Petrovic Fabijan A, Lin RCY, Ho J, Maddocks S, Ben Zakour NL, Iredell JR	2020	Safety of bacteriophage therapy in severe *Staphylococcus aureus* infection	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA541589	NCBI BioProject, PRJNA541589

References

Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R. 2017. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**:446–451. DOI: https://doi.org/10.1038/nature22364

Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. 2015. Analysis of Bottlenecks in Experimental Models of Infection. *PLOS Pathogens* **11**:e1004823. DOI: https://doi.org/10.1371/journal.ppat.1004823

Altman DR, Sullivan MJ, Chacko KI, Balasubramanian D, Pak TR, Sause WE, Kumar K, Sebra R, Deikus G, Attie O, Rose H, Lewis M, Fulmer Y, Bashir A, Kasarskis A, Schadt EE, Richardson AR, Torres VJ, Shopsin B, van Bakel H. 2018. Genome Plasticity of agr-Defective Staphylococcus aureus during Clinical Infection. *Infection and Immunity* **86**:e00331-18. DOI: https://doi.org/10.1128/IAI.00331-18, PMID: 30061376

Aseev LV, Boni IV. 2011. Extraribosomal functions of bacterial ribosomal proteins. *Molekuliarnaia Biologiia* **45**:805–816. DOI: https://doi.org/10.1134/S0026893311050025, PMID: 22393776

Austin CM, Garabaglu S, Krute CN, Ridder MJ, Seawell NA, Markiewicz MA. 2019. Contribution of YjbIH to Virulence Factor Expression and Host Colonization in Staphylococcus aureus. *Infection and Immunity* **87**:e00155-19. DOI: https://doi.org/10.1128/IAI.00155-19

Azarian T, Ridgway JP, Yin Z, David MZ. 2019. Long-Term Intrahost Evolution of Methicillin Resistant Staphylococcus aureus Among Cystic Fibrosis Patients With Respiratory Carriage. *Frontiers in Genetics* **10**:546. DOI: https://doi.org/10.3389/fgene.2019.00546

Bailey SF, Blanquart F, Bataillon T, Kassen R. 2017. What drives parallel evolution?: How population size and mutational variation contribute to repeated evolution. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* **39**:1–9. DOI: https://doi.org/10.1002/bies.201600176, PMID: 27859467

Basuino L, Jousselin A, Alexander JAN, Strynadka NCJ, Pinho MG, Chambers HF. 2018. PBP4 activity and its overexpression are necessary for PBP4-mediated high-level β-lactam resistance. *The Journal of Antimicrobial Chemotherapy* **73**:1177–1180. DOI: https://doi.org/10.1093/jac/dkx531

Batut B, Knibbe C, Marais G, Daubin V. 2014. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nature Reviews. Microbiology* **12**:841–850. DOI: https://doi.org/10.1038/nrmicro3331

Behe MJ. 2010. Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution." *The Quarterly Review of Biology* **85**:419–445. DOI: https://doi.org/10.1086/656902, PMID: 21243963

Benoit JB, Frank DN, Bessesen MT. 2018. Genomic evolution of Staphylococcus aureus isolates colonizing the nares and progressing to bacteremia. *PLOS ONE* **13**:e0195860. DOI: https://doi.org/10.1371/journal.pone.0195860

Both A, Huang J, Qi M, Lausmann C, Weißelberg S, Büttner H. 2021. Distinct clonal lineages and within-host diversification shape invasive Staphylococcus epidermidis populations. *PLOS Pathogens* **17**:e1009304. DOI: https://doi.org/10.1371/journal.ppat.1009304

Bryant JM, Brown KP, Burbaud S, Everall I, Belardinelli JM, Rodriguez-Rincon D, Grogono DM, Peterson CM, Verma D, Evans IE, Ruis C, Weimann A, Arora D, Malhotra S, Bannerman B, Passemar C, Templeton K, MacGregor G, Jiwa K, Fisher AJ, et al. 2021. Stepwise pathogenic evolution of Mycobacterium abscessus. *Science (New York, N.Y.)* **372**:eabb8699. DOI: https://doi.org/10.1126/science.abb8699, PMID: 33926925

Brynildsrud O, Snipen LG, Bohlin J. 2015. CNOGpro: detection and quantification of CNVs in prokaryotic whole-genome sequencing data. *Bioinformatics (Oxford, England)* **31**:1708–1715. DOI: https://doi.org/10.1093/bioinformatics/btv070

Burd EM, Alam MT, Passalacqua KD, Kalokhe AS, Eaton ME, Satola SW, Kraft CS, Read TD. 2014. Development of oxacillin resistance in a patient with recurrent Staphylococcus aureus bacteremia. *Journal of Clinical Microbiology* **52**:3114–3117. DOI: https://doi.org/10.1128/JCM.00615-14, PMID: 24850355

CameronDR, WardDV, KostouliasX, HowdenBP, Moellering RC J, EliopoulosGM. 2012. Serine/threonine phosphatase Stp1 contributes to reduced susceptibility to vancomycin and virulence in Staphylococcus aureus. *The Journal of Infectious Diseases* **205**:1677–1687. DOI: https://doi.org/10.1093/infdis/jis252

Chaguza C, Senghore M, Bojang E, Gladstone RA, LoSW, Tientcheu PE. 2020. Within-host microevolution of Streptococcus pneumoniae is rapid and adaptive during natural colonisation. *Nature Communications* **11**:3442. DOI: https://doi.org/10.1038/s41467-020-17327-w

Chilambi GS, Nordstrom HR, Evans DR, Ferrolino JA, Hayden RT, Marón GM, Vo AN, Gilmore MS, Wolf J, Rosch JW, Van Tyne D. 2020. Evolution of vancomycin-resistant Enterococcus faecium during colonization and infection in immunocompromised pediatric patients. *PNAS* **117**:11703–11714. DOI: https://doi.org/10.1073/pnas.1917130117, PMID: 32393645

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE* **7**:e46688. DOI: https://doi.org/10.1371/journal.pone.0046688, PMID: 23056405

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**:80–92. DOI: https://doi.org/10.4161/fly.19695

Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Research* **38**:D396-D400. DOI: https://doi.org/10.1093/nar/gkp919, PMID: 19906701

Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016. Within-host evolution of bacterial pathogens. *Nature Reviews. Microbiology* **14**:150–162. DOI: https://doi.org/10.1038/nrmicro2015.13

Engman J, Rogstam A, Frees D, Ingmer H, von Wachenfeldt C. 2012. The YjbH adaptor protein enhances proteolysis of the transcriptional regulator Spx in Staphylococcus aureus. *Journal of Bacteriology* **194**:1186–1194. DOI: https://doi.org/10.1128/JB.06414-11

Ernst CM, Peschel A. 2019. MprF-mediated daptomycin resistance. *International Journal of Medical Microbiology* **309**:359–363. DOI: https://doi.org/10.1016/j.ijmm.2019.05.010

Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu C-H, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W. 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrobial Agents and Chemotherapy* **63**:e00483-19. DOI: https://doi.org/10.1128/AAC.00483-19, PMID: 31427293

Fuchs S, Mehlan H, Bernhardt J, Hennig A, Michalik S, Surmann K, Pané-Farré J, Giese A, Weiss S, Backert L, Herbig A, Nieselt K, Hecker M, Völker U, Mäder U. 2018. AureoWiki- The repository of the Staphylococcus aureus research and annotation community. *International Journal of Medical Microbiology* **308**:558–568. DOI: https://doi.org/10.1016/j.ijmm.2017.11.011, PMID: 29198880

Gao W, Cameron DR, Davies JK, Kostoulias X, Stepnell J, Tuck KL, Yeaman MR, Peleg AY, Stinear TP, Howden BP. 2013. The RpoB H$_{481}$Y rifampicin resistance mutation and an active stringent response reduce virulence and increase resistance to innate immune responses in Staphylococcus aureus. *The Journal of Infectious Diseases* **207**:929–939. DOI: https://doi.org/10.1093/infdis/jis772, PMID: 23255563

Gao W, Monk IR, Tobias NJ, Gladman SL, Seemann T, Stinear TP, Howden BP. 2015. Large tandem chromosome expansions facilitate niche adaptation during persistent infection with drug-resistant Staphylococcus aureus. *Microbial Genomics* **1**:e000026. DOI: https://doi.org/10.1099/mgen.0.000026, PMID: 28348811

Gatt YE, Margalit H. 2021. Common Adaptive Strategies Underlie Within-Host Evolution of Bacterial Pathogens. *Molecular Biology and Evolution* **38**:1101–1121. DOI: https://doi.org/10.1093/molbev/msaa278, PMID: 33118035

Gerasimavicius L, Livesey BJ, Marsh JA. 2021. Loss-of-Function, Gain-of-Function and Dominant-Negative Mutations Have Profoundly Different Effects on Protein Structure: Implications for Variant Effect Prediction. *bioRxiv*. DOI: https://doi.org/10.1101/2021.10.23.465554

Giulieri SG, Baines SL, Guerillot R, Seemann T, Gonçalves da Silva A, Schultz M. 2018. Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent Staphylococcus aureus bacteraemia. *Genome Medicine* **10**:65. DOI: https://doi.org/10.1186/s13073-018-0574-x

Giulieri SG, Guérillot R, Kwong JC, Monk IR, Hayes AS, Daniel D, Baines S, Sherry NL, Holmes NE, Ward P, Gao W, Seemann T, Stinear TP, Howden BP. 2020. Comprehensive Genomic Investigation of Adaptive Mutations Driving the Low-Level Oxacillin Resistance Phenotype in Staphylococcus aureus. *MBio* **11**:e02882-20. DOI: https://doi.org/10.1128/mBio.02882-20, PMID: 33293382

Göhring N, Fedtke I, Xia G, Jorge AM, Pinho MG, Bertsche U. 2011. New role of the disulfide stress effector YjbH in β-lactam susceptibility of Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* **55**:5452–5458. DOI: https://doi.org/10.1128/AAC.00286-11

Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H. 2013. Within-host evolution of Staphylococcus aureus during asymptomatic carriage. *PLOS ONE* **8**:e61319. DOI: https://doi.org/10.1371/journal.pone.0061319

Goncheva MI, Flannagan RS, Sterling BE, Laakso HA, Friedrich NC, Kaiser JC. 2019. Stress-induced inactivation of the Staphylococcus aureus purine biosynthesis repressor leads to hypervirulence. *Nature Communications* **10**:775. DOI: https://doi.org/10.1038/s41467-019-08724-x

Goncheva MI, Flannagan RS, Heinrichs DE. 2020. De Novo Purine Biosynthesis Is Required for Intracellular Growth of Staphylococcus aureus and for the Hypervirulence Phenotype of a purR Mutant. *Infection and Immunity* **88**:e00104-20. DOI: https://doi.org/10.1128/IAI.00104-20

Griffith DM, Veech JA, Marsh CJ. 2016. cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software, Code Snippets* **69**:1–17. DOI: https://doi.org/10.18637/jss.v069.c02

Guérillot R, Gonçalves da Silva A, Monk I, Giulieri S, Tomita T, Alison E, Porter J, Pidot S, Gao W, Peleg AY, Seemann T, Stinear TP, Howden BP. 2018. Convergent Evolution Driven by Rifampin Exacerbates the Global Burden of Drug-Resistant Staphylococcus aureus. *MSphere* **3**:e00550-17. DOI: https://doi.org/10.1128/mSphere.00550-17, PMID: 29404415

Guerillot R, Li L, Baines S, Howden B, Schultz MB, Seemann T. 2018. Comprehensive antibiotic-linked mutation assessment by resistance mutation sequencing (RM-seq. *Genome Medicine* **10**:63. DOI: https://doi.org/10.1186/s13073-018-0572-z

Guérillot R, Kostoulias X, Donovan L, Li L, Carter GP, Hachani A, Vandelannoote K, Giulieri S, Monk IR, Kunimoto M, Starrs L, Burgio G, Seemann T, Peleg AY, Stinear TP, Howden BP. 2019. Unstable chromosome rearrangements in Staphylococcus aureus cause phenotype switching associated with persistent infections. *PNAS* **116**:20135–20140. DOI: https://doi.org/10.1073/pnas.1904861116, PMID: 31527262

Hall MD, Holden MT, Srisomang P, Mahavanakul W, Wuthiekanun V, Limmathurotsakul D, Fountain K, Parkhill J, Nickerson EK, Peacock SJ, Fraser C. 2019. Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *eLife* **8**:e46402. DOI: https://doi.org/10.7554/eLife.46402, PMID: 31591959

Harkins CP, Pettigrew KA, Oravcova K, Gardner J, Hearn RMR, Rice D. 2018. The Microevolution and Epidemiology of Staphylococcus aureus Colonization during Atopic Eczema Disease Flare. *The Journal of Investigative Dermatology* **138**:336–343. DOI: https://doi.org/10.1016/j.jid.2017.09.023

Hawkey J, Monk JM, Billman-Jacobe H, Palsson B, Holt KE. 2020. Impact of insertion sequences on convergent evolution of Shigella species. *PLOS Genetics* **16**:e1008931. DOI: https://doi.org/10.1371/journal.pgen.1008931

Hood MI, Skaar EP. 2012. Nutritional immunity: transition metals at the pathogen-host interface. *Nature Reviews. Microbiology* **10**:525–537. DOI: https://doi.org/10.1038/nrmicro2836, PMID: 22796883

Howden BP, Johnson PD, Ward PB, Stinear TP, Davies JK. 2006. Isolates with low-level vancomycin resistance associated with persistent methicillin-resistant Staphylococcus aureus bacteremia. *Antimicrobial Agents and Chemotherapy* **50**:3039–3047. DOI: https://doi.org/10.1128/AAC.00422-06

Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF. 2011. Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. *PLOS Pathogens* **7**:e1002359. DOI: https://doi.org/10.1371/journal.ppat.1002359

Ji S, Jiang S, Wei X, Sun L, Wang H, Zhao F, Chen Y, Yu Y. 2020. In-Host Evolution of Daptomycin Resistance and Heteroresistance in Methicillin-Resistant Staphylococcus aureus Strains From Three Endocarditis Patients. *The Journal of Infectious Diseases* **221**:S243–S252. DOI: https://doi.org/10.1093/infdis/jiz571, PMID: 32176794

Jiang JH, Dexter C, Cameron DR, Monk IR, Baines SL, Abbott IJ. 2019. Evolution of Daptomycin Resistance in Coagulase-Negative Staphylococci Involves Mutations of the Essential Two-Component Regulator WalKR. *Antimicrobial Agents and Chemotherapy* **63**:e01926-18. DOI: https://doi.org/10.1128/AAC.01926-18

Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, Reinhardt R, Correa P, Meyer TF, Josenhans C, Falush D, Suerbaum S. 2011. Helicobacter pylori genome evolution during human infection. *PNAS* **108**:5033–5038. DOI: https://doi.org/10.1073/pnas.1018444108, PMID: 21383187

Key FM, Khadka VD, Romo-González C, Blake KJ, Deng L, Lynn TC, Lee JC, Chiu IM, García-Romero MT, Lieberman TD. 2021. On-Person Adaptive Evolution of *Staphylococcus aureus* during Treatment for Atopic Dermatitis. *bioRxiv*. DOI: https://doi.org/10.1101/2021.03.24.436824

Khademi SMH, Sazinas P, Jelsbak L. 2019. Within-Host Adaptation Mediated by Intergenic Evolution in Pseudomonas aeruginosa. *Genome Biology and Evolution* **11**:1385–1397. DOI: https://doi.org/10.1093/gbe/evz083, PMID: 30980662

Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C. 2016. Emergence of host-adapted Salmonella Enteritidis through rapid evolution in an immunocompromised host. *Nature Microbiology* **1**:e15023. DOI: https://doi.org/10.1038/nmicrobiol.2015.23

Krismer B, Weidenmaier C, Zipperer A, Peschel A. 2017. The commensal lifestyle of Staphylococcus aureus and its interactions with the nasal microbiota. *Nature Reviews. Microbiology* **15**:675–687. DOI: https://doi.org/10.1038/nrmicro.2017.104

Kuehl R, Morata L, Meylan S, Mensa J, Soriano A. 2020. When antibiotics fail: a clinical and microbiological perspective on antibiotic tolerance and persistence of Staphylococcus aureus. *The Journal of Antimicrobial Chemotherapy* **75**:1071–1086. DOI: https://doi.org/10.1093/jac/dkz559

Kuroda M, Sekizuka T, Matsui H, Ohsuga J, Ohshima T, Hanaki H. 2019. IS256-Mediated Overexpression of the WalKR Two-Component System Regulon Contributes to Reduced Vancomycin Susceptibility in a Staphylococcus aureus Clinical Isolate. *Frontiers in Microbiology* **10**:1882. DOI: https://doi.org/10.3389/fmicb.2019.01882, PMID: 31474962

Laabei M, Uhlemann AC, Lowy FD, Austin ED, Yokoyama M, Ouadi K. 2015. Evolutionary Trade-Offs Underlie the Multi-faceted Virulence of Staphylococcus aureus. *PLOS Biology* **13**:e1002229. DOI: https://doi.org/10.1371/journal.pbio.1002229

Langhanki L, Berger P, Treffon J, Catania F, Kahl BC, Mellmann A. 2018. In vivo competition and horizontal gene transfer among distinct Staphylococcus aureus lineages as major drivers for adaptational changes during long-term persistence in humans. *BMC Microbiology* **18**:152. DOI: https://doi.org/10.1186/s12866-018-1308-3

Lees JA, Croucher NJ, Goldblatt D, Nosten F, Parkhill J, Turner C, Turner P, Bentley SD. 2017a. Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. *eLife* **6**:e26255. DOI: https://doi.org/10.7554/eLife.26255, PMID: 28742023

Lees JA, Kremer PHC, Manso AS, Croucher NJ, Ferwerda B, Serón MV, Oggioni MR, Parkhill J, Brouwer MC, van der Ende A, van de Beek D, Bentley SD. 2017b. Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. *Microbial Genomics* **3**:e000103. DOI: https://doi.org/10.1099/mgen.0.000103, PMID: 28348877

Lemoine F, Gascuel O. 2021. Gotree/Goalign: Toolkit and Go API to Facilitate the Development of Phylogenetic Workflows. *bioRxiv*. DOI: https://doi.org/10.1101/2021.06.09.447704

Levin-Reisman I, Brauner A, Ronin I, Balaban NQ. 2019. Epistasis between antibiotic tolerance, persistence, and resistance mutations. *PNAS* **116**:14734–14739. DOI: https://doi.org/10.1073/pnas.1906169116, PMID: 31262806

Li H. 2013. Aligning Sequence Reads, Clone Sequences and Assembly CContigs with BWA-MEM. [arXiv]. https://doi.org/10.48550/arXiv.1303.3997

Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T. 2016. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. *Nature Medicine* **22**:1470–1474. DOI: https://doi.org/10.1038/nm.4205

Liu J, Gefen O, Ronin I, Bar-Meir M, Balaban NQ. 2020. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science (New York, N.Y.)* **367**:200–204. DOI: https://doi.org/10.1126/science.aay3041, PMID: 31919223

Long DR, Wolter DJ, Lee M, Precit M, McLean K, Holmes E. 2020. Polyclonality, Shared Strains, and Convergent Evolution in Chronic CF S. aureus Airway Infection. *American Journal of Respiratory and Critical Care Medicine* **203**:1127–1137. DOI: https://doi.org/10.1164/rccm.202003-0735OC

Lopatkin AJ, Bening SC, Manson AL, Stokes JM, Kohanski MA, Badran AH. 2021. Clinically relevant mutations in core metabolic genes confer antibiotic resistance. *Science (New York, N.Y.)* **371**:eaba0862. DOI: https://doi.org/10.1126/science.aba0862

Loss G, Simoes PM, Valour F, Cortes MF, Gonzaga L, Bergot M. 2019. Staphylococcus aureus Small Colony Variants (SCVs): News From a Chronic Prosthetic Joint Infection. *Frontiers in Cellular and Infection Microbiology* **9**:363. DOI: https://doi.org/10.3389/fcimb.2019.00363

Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J. Journal of Open Source Software* **6**:3139. DOI: https://doi.org/10.21105/joss.03139

Machado H, Seif Y, Sakoulas G, Olson CA, Hefner Y, Anand A. 2021. Environmental conditions dictate differential evolution of vancomycin resistance in Staphylococcus aureus. *Communications Biology* **4**:e793. DOI: https://doi.org/10.1038/s42003-021-02339-z

Mäder U, Nicolas P, Depke M, Pané-Farré J, Debarbouille M, van der Kooi-Pol MM, Guérin C, Dérozier S, Hiron A, Jarmer H, Leduc A, Michalik S, Reilman E, Schaffer M, Schmidt F, Bessières P, Noirot P, Hecker M, Msadek T, Völker U, et al. 2016. Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. *PLOS Genetics* **12**:e1005962. DOI: https://doi.org/10.1371/journal.pgen.1005962, PMID: 27035918

Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**:1029–1041. DOI: https://doi.org/10.1016/j.cell.2017.09.042

Marvig RL, Sommer LM, Molin S, Johansen HK. 2015. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nature Genetics* **47**:57–64. DOI: https://doi.org/10.1038/ng.3148

McEvoy CRE, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP. 2013. Decreased vancomycin susceptibility in Staphylococcus aureus caused by IS256 tempering of WalKR expression. *Antimicrobial Agents and Chemotherapy* **57**:3240–3249. DOI: https://doi.org/10.1128/AAC.00279-13, PMID: 23629723

McVicker G, Prajsnar TK, Williams A, Wagner NL, Boots M, Renshaw SA. 2014. Clonal expansion during Staphylococcus aureus infection dynamics reveals the effect of antibiotic intervention. *PLOS Pathogens* **10**:e1003959. DOI: https://doi.org/10.1371/journal.ppat.1003959

Miller CR, Dey S, Smolenski PD, Kulkarni PS, Monk JM, Szubin R, Sakoulas G, Berti AD. 2020. Distinct Subpopulations of Intravalvular Methicillin-Resistant Staphylococcus aureus with Variable Susceptibility to Daptomycin in Tricuspid Valve Endocarditis. *Antimicrobial Agents and Chemotherapy* **64**:e01593-19. DOI: https://doi.org/10.1128/AAC.01593-19, PMID: 31932377

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**:1530–1534. DOI: https://doi.org/10.1093/molbev/msaa015, PMID: 32011700

Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ. 2015. Complete Bypass of Restriction Systems for Major Staphylococcus aureus Lineages. *MBio* **6**:e00308-15. DOI: https://doi.org/10.1128/mBio.00308-15, PMID: 26015493

Monk IR, Shaikh N, Begg SL, Gajdiss M, Sharkey LKR, Lee JYH, Pidot SJ, Seemann T, Kuiper M, Winnen B, Hvorup R, Collins BM, Bierbaum G, Udagedara SR, Morey JR, Pulyani N, Howden BP, Maher MJ, McDevitt CA, King GF, et al. 2019. Zinc-binding to the cytoplasmic PAS domain regulates the essential WalK histidine kinase of Staphylococcus aureus. *Nature Communications* **10**:3067. DOI: https://doi.org/10.1038/s41467-019-10932-4, PMID: 31296851

Morvan C, Halpern D, Kénanian G, Hays C, Anba-Mondoloni J, Brinster S. 2016. Environmental fatty acids enable emergence of infectious Staphylococcus aureus resistant to FASII-targeted antimicrobials. *Nature Communications* **7**:12944. DOI: https://doi.org/10.1038/ncomms12944

Novick RP, Geisinger E. 2008. Quorum sensing in staphylococci. *Annual Review of Genetics* **42**:541–564. DOI: https://doi.org/10.1146/annurev.genet.42.110807.091640

Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* **2**:e000056. DOI: https://doi.org/10.1099/mgen.0.000056, PMID: 28348851

Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, Holden MTG. 2015. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nature Communications* **6**:6560. DOI: https://doi.org/10.1038/ncomms7560

Paudel A, Panthee S, Hamamoto H, Grunert T, Sekimizu K. 2021. YjbH regulates virulence genes expression and oxidative stress resistance in Staphylococcus aureus. *Virulence* **12**:470–480. DOI: https://doi.org/10.1080/21505594.2021.1875683, PMID: 33487122

Peleg AY, Miyakis S, Ward DV, Earl AM, Rubio A, Cameron DR, Pillai S, Moellering RC, Eliopoulos GM. 2012. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of Staphylococcus aureus. *PLOS ONE* **7**:e28316. DOI: https://doi.org/10.1371/journal.pone.0028316, PMID: 22238576

Petrovic Fabijan A, Lin RCY, Ho J, Maddocks S, Ben Zakour NL, Iredell JR, Westmead Bacteriophage Therapy Team. 2020. Safety of bacteriophage therapy in severe Staphylococcus aureus infection. *Nature Microbiology* **5**:465–472. DOI: https://doi.org/10.1038/s41564-019-0634-z, PMID: 32066959

Phaneuf PV, Yurkovich JT, Heckmann D, Wu M, Sandberg TE, King ZA. 2020. Causal mutations from adaptive laboratory evolution are outlined by multiple scales of genome annotations and condition-specificity. *BMC Genomics* **21**:514. DOI: https://doi.org/10.1186/s12864-020-06920-4

Phillips PC. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**:855–867. DOI: https://doi.org/10.1038/nrg2452

Prados J, Linder P, Redder P. 2016. TSS-EMOTE, a refined protocol for a more complete and less biased global mapping of transcription start sites in bacterial pathogens. *BMC Genomics* **17**:849. DOI: https://doi.org/10.1186/s12864-016-3211-3

Price MN, Arkin AP, Langille MGI. 2017. PaperBLAST: Text Mining Papers for Information about Homologs. *MSystems* **2**:e00039-17. DOI: https://doi.org/10.1128/mSystems.00039-17

Proctor RA. 2019. Immunity to Staphylococcus aureus: Implications for Vaccine Development. *Microbiology Spectrum* **7**, e0037-2018. DOI: https://doi.org/10.1128/microbiolspec.GPP3-0037-2018, PMID: 31298209

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**:841–842. DOI: https://doi.org/10.1093/bioinformatics/btq033

Renzoni A, Andrey DO, Jousselin A, Barras C, Monod A, Vaudaux P. 2011. Whole genome sequencing and complete genetic analysis reveals novel pathways to glycopeptide resistance in Staphylococcus aureus. *PLOS ONE* **6**:e21577. DOI: https://doi.org/10.1371/journal.pone.0021577

Rishishwar L, Kraft CS, Jordan IK. 2016. Population Genomics of Reduced Vancomycin Susceptibility in Staphylococcus aureus. *MSphere* **1**:e00094. DOI: https://doi.org/10.1128/mSphere.00094-16

Rouard C, Garnier F, Leraut J, Lepainteur M, Rahajamananav L, Languepin J, Ploy M-C, Bourgeois-Nicolaos N, Doucet-Populaire F. 2018. Emergence and Within-Host Genetic Evolution of Methicillin-Resistant Staphylococcus aureus Resistant to Linezolid in a Cystic Fibrosis Patient. *Antimicrobial Agents and Chemotherapy* **62**:e00720-18. DOI: https://doi.org/10.1128/AAC.00720-18, PMID: 30275089

Sause WE, Balasubramanian D, Irnov I, Copin R, Sullivan MJ, Sommerfield A. 2019. The purine biosynthesis regulator PurR moonlights as a virulence regulator in Staphylococcus aureus. *PNAS* **116**:13563–13572. DOI: https://doi.org/10.1073/pnas.1904280116

Schreiber F, Szekat C, Josten M, Sahl HG, Bierbaum G. 2013. Antibiotic-induced autoactivation of IS256 in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* **57**:6381–6384. DOI: https://doi.org/10.1128/AAC.01585-13, PMID: 24080654

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)* **30**:2068–2069. DOI: https://doi.org/10.1093/bioinformatics/btu153

Seemann T. 2022a. ABRicate. 955d402. GitHub. https://github.com/tseemann/abricate

Seemann T. 2022b. mlst. 09851e9. GitHub. https://github.com/tseemann/mlst

Seemann T. 2022c. Shovill. f64d19e. GitHub. https://github.com/tseemann/shovill

Seemann T. 2022d. snippy. 2af674c. Github. https://github.com/tseemann/snippy

Shopsin B, Drlica-Wagner A, Mathema B, Adhikari RP, Kreiswirth BN, Novick RP. 2008. Prevalence of agr dysfunction among colonizing Staphylococcus aureus strains. *The Journal of Infectious Diseases* **198**:1171–1174. DOI: https://doi.org/10.1086/592051, PMID: 18752431

Skwark MJ, Croucher NJ, Puranen S, Chewapreecha C, Pesonen M. 2017. Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLOS Genetics* **13**:e1006508. DOI: https://doi.org/10.1371/journal.pgen.1006508

Smyth DS, Kafer JM, Wasserman GA, Velickovic L, Mathema B, Holzman RS. 2012. Nasal carriage as a source of agr-defective Staphylococcus aureus bacteremia. *The Journal of Infectious Diseases* **206**:1168–1177. DOI: https://doi.org/10.1093/infdis/jis483

Stinear TP, Seemann T, Pidot S, Frigui W, Reysset G, Garnier T. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. *Genome Research* **17**:192–200. DOI: https://doi.org/10.1101/gr.5942807

Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Molecular Biology and Evolution* **28**:63–70. DOI: https://doi.org/10.1093/molbev/msq249, PMID: 20837603

Suligoy CM, Lattar SM, Noto Llana M, Gonzalez CD, Alvarez LP, Robinson DA. 2018. Mutation of Agr Is Associated with the Adaptation of Staphylococcus aureus to the Host during Chronic Osteomyelitis. *Frontiers in Cellular and Infection Microbiology* **8**:18. DOI: https://doi.org/10.3389/fcimb.2018.00018

Suligoy CM, Díaz RE, Gehrke AK, Ring N, Yebra G, Alves J. 2020. Acapsular Staphylococcus aureus with a non-functional agr regains capsule expression after passage through the bloodstream in a bacteremia mouse model. *Scientific Reports* **10**:14108. DOI: https://doi.org/10.1038/s41598-020-70671-1

Surewaard BGJ, Deniset JF, Zemp FJ, Amrein M, Otto M, Conly J, Omri A, Yates RM, Kubes P. 2016. Identification and treatment of the Staphylococcus aureus reservoir in vivo. *The Journal of Experimental Medicine* **213**:1141–1151. DOI: https://doi.org/10.1084/jem.20160334, PMID: 27325887

Tan X, Coureuil M, Ramond E, Euphrasie D, Dupuis M, Tros F, Meyer J, Nemazanyy I, Chhuon C, Guerrera IC, Ferroni A, Sermet-Gaudelus I, Nassif X, Charbit A, Jamet A. 2019. Chronic Staphylococcus aureus Lung Infection Correlates With Proteogenomic and Metabolic Adaptations Leading to an Increased Intracellular Persistence. *Clinical Infectious Diseases* **69**:1937–1945. DOI: https://doi.org/10.1093/cid/ciz106, PMID: 30753350

Thwaites GE, Gant V. 2011. Are bloodstream leukocytes Trojan Horses for the metastasis of Staphylococcus aureus? *Nature Reviews. Microbiology* **9**:215–222. DOI: https://doi.org/10.1038/nrmicro2508

Toft C, Andersson SG. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews. Genetics* **11**:465–475. DOI: https://doi.org/10.1038/nrg2798

Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. 2015. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical Microbiology Reviews* **28**:603–661. DOI: https://doi.org/10.1128/CMR.00134-14, PMID: 26016486

Tong SYC, Lye DC, Yahav D, Sud A, Robinson JO, Nelson J, Archuleta S, Roberts MA, Cass A, Paterson DL, Foo H, Paul M, Guy SD, Tramontana AR, Walls GB, McBride S, Bak N, Ghosh N, Rogers BA, Ralph AP, et al. 2020. Effect of Vancomycin or Daptomycin With vs Without an Antistaphylococcal β-Lactam on Mortality, Bacteremia, Relapse, or Treatment Failure in Patients With MRSA Bacteremia: A Randomized Clinical Trial. *JAMA* **323**:527–537. DOI: https://doi.org/10.1001/jama.2020.0103, PMID: 32044943

Törönen P, Medlar A, Holm L. 2018. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research* **46**:W84–W88. DOI: https://doi.org/10.1093/nar/gky350

Trong HNG, Prunier AL, Leclercq R. 2005. Hypermutable and fluoroquinolone-resistant clinical isolates of Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* **49**:2098–2101. DOI: https://doi.org/10.1128/AAC.49.5.2098-2101.2005, PMID: 15855537

Trouillet-Assant S, Lelièvre L, Martins-Simões P, Gonzaga L, Tasse J, Valour F, Rasigade JP, Vandenesch F, Muniz Guedes RL, Ribeiro de Vasconcelos AT, Caillon J, Lustig S, Ferry T, Jacqueline C, Loss de Morais G, Laurent F. 2016. Adaptive processes of Staphylococcus aureus isolates during the progression from acute to chronic bone and joint infections in patients. *Cellular Microbiology* **18**:1405–1414. DOI: https://doi.org/10.1111/cmi.12582, PMID: 26918656

Van Tyne D, Manson AL, Huycke MM, Karanicolas J, Earl AM, Gilmore MS. 2019. Impact of antibiotic treatment and host innate immune pressure on enterococcal adaptation in the human bloodstream. *Science Translational Medicine* **11**:eaat8418. DOI: https://doi.org/10.1126/scitranslmed.aat8418, PMID: 30971455

Vargas R, Freschi L, Marin M, Epperson LE, Smith M, Oussenko I. 2021. In-host population dynamics of Mycobacterium tuberculosis complex during active disease. *eLife* **10**:e61805. DOI: https://doi.org/10.7554/eLife.61805

Wadsworth CB, Arnold BJ, Sater MRA, Grad YH. 2018. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in Neisseria gonorrhoeae. *MBio* **9**:e01419-18. DOI: https://doi.org/10.1128/mBio.01419-18

Wang Y, Bojer MS, George SE, Wang Z, Jensen PR, Wolz C, Ingmer H. 2018. Inactivation of TCA cycle enhances Staphylococcus aureus persister cell formation in stationary phase. *Scientific Reports* **8**:10849. DOI: https://doi.org/10.1038/s41598-018-29123-0, PMID: 30022089

Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA. 2005. The role of nasal carriage in Staphylococcus aureus infections. *The Lancet. Infectious Diseases* **5**:751–762. DOI: https://doi.org/10.1016/S1473-3099(05)70295-4

Wheatley R, Diaz Caballero J, Kapel N, de Winter FHR, Jangir P, Quinn A, Del Barrio-Tofiño E, López-Causapé C, Hedge J, Torrens G, Van der Schalk T, Xavier BB, Fernández-Cuenca F, Arenzana A, Recanatini C, Timbermont L, Sifakis F, Ruzin A, Ali O, Lammens C, et al. 2021. Rapid evolution and host immunity drive the rise and fall of carbapenem resistance during an acute Pseudomonas aeruginosa infection. *Nature Communications* **12**:2460. DOI: https://doi.org/10.1038/s41467-021-22814-9, PMID: 33911082

Wheeler NE, Gardner PP, Barquist L. 2018. Machine learning identifies signatures of host adaptation in the bacterial pathogen Salmonella enterica. *PLOS Genetics* **14**:e1007333. DOI: https://doi.org/10.1371/journal.pgen.1007333

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**:257. DOI: https://doi.org/10.1186/s13059-019-1891-0, PMID: 31779668

Wuthrich D, Cuenod A, Hinic V, Morgenstern M, Khanna N, Egli A. 2019. Genomic characterization of inpatient evolution of MRSA resistant to daptomycin, vancomycin and ceftaroline. *The Journal of Antimicrobial Chemotherapy* **74**:1452–1454. DOI: https://doi.org/10.1093/jac/dkz003

Yokoyama M, Stevens E, Laabei M, Bacon L, Heesom K, Bayliss S. 2018. Epistasis analysis uncovers hidden antibiotic resistance-associated fitness costs hampering the evolution of MRSA. *Genome Biology* **19**:94. DOI: https://doi.org/10.1186/s13059-018-1469-2

Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CLC, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, et al. 2012. Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. *PNAS* **109**:4550–4555. DOI: https://doi.org/10.1073/pnas.1113219109, PMID: 22393007

Young BC, Gordon NC, Cole K, Price JR, Liu E. 2017. Severe infections emerge from commensal bacteria by adaptive evolution. *eLife* **6**:e30637. DOI: https://doi.org/10.7554/eLife.30637

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS : A Journal of Integrative Biology* **16**:284–287. DOI: https://doi.org/10.1089/omi.2011.0118, PMID: 22455463