







The mutational landscape of *Staphylococcus aureus* during colonisation

Received: 2 August 2024

Accepted: 2 December 2024

Published online: 13 January 2025

 Check for updates

Francesc Coll^{1,2,3,13} , Beth Blane^{4,13}, Katherine L. Bellis^{3,4,13}, Marta Matuszewska^{4,5} , Toska Wonfor^{6,7}, Dorota Jamroz³, Michelle S. Toleman⁴, Joan A. Geoghegan^{6,7,8} , Julian Parkhill⁵, Ruth C. Massey^{9,10,11} , Sharon J. Peacock^{4,14}  & Ewan M. Harrison^{3,4,12,14} 

Staphylococcus aureus is an important human pathogen and a commensal of the human nose and skin. Survival and persistence during colonisation are likely major drivers of *S. aureus* evolution. Here we applied a genome-wide mutation enrichment approach to a genomic dataset of 3060 *S. aureus* colonization isolates from 791 individuals. Despite limited within-host genetic diversity, we observed an excess of protein-altering mutations in metabolic genes, in regulators of quorum-sensing (*agrA* and *agrC*) and in known antibiotic targets (*fusA*, *pbp2*, *dfrA* and *ileS*). We demonstrated the phenotypic effect of multiple adaptive mutations in vitro, including changes in haemolytic activity, antibiotic susceptibility, and metabolite utilisation. Nitrogen metabolism showed the strongest evidence of adaptation, with the assimilatory nitrite reductase (*nasD*) and urease (*ureG*) showing the highest mutational enrichment. We identified a *nasD* natural mutant with enhanced growth under urea as the sole nitrogen source. Inclusion of 4090 additional isolate genomes from 731 individuals revealed eight more genes including *sasA/sraP*, *darA/pstA*, and *rsbU* with signals of adaptive variation that warrant further characterisation. Our study provides a comprehensive picture of the heterogeneity of *S. aureus* adaptive changes during colonisation, and a robust methodological approach applicable to study *in host* adaptive evolution in other bacterial pathogens.

Staphylococcus aureus is an important pathogen but also a commensal bacterium and part of the human microbiota. The anterior nares (lower nostrils) are the primary reservoir of *S. aureus* in humans, although the bacterium may colonise other body sites including the skin, pharynx,

axillae, perineum and the intestine^{1,2}. Despite being a commensal, when the epithelial barrier breaks or the immune system becomes compromised, *S. aureus* can cause a variety of infections, ranging from superficial skin and soft-tissue infections to life-threatening invasive

¹Applied Microbial Genomics Unit, Department of Molecular Basis of Disease, Institute of Biomedicine of Valencia (IBV-CSIC), Valencia, Spain. ²Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. ³Parasites & Microbes Programme, Wellcome Sanger Institute, Hinxton, UK. ⁴Department of Medicine, University of Cambridge, Cambridge, UK. ⁵Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. ⁶Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. ⁷Department of Microbes, Infection & Microbiomes, College of Medicine & Health, University of Birmingham, Birmingham, UK. ⁸Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin, Ireland. ⁹School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK. ¹⁰School of Microbiology, University College Cork, Cork, Ireland. ¹¹APC Microbiome Ireland, University College Cork, Cork, Ireland. ¹²Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ¹³These authors contributed equally: Francesc Coll, Beth Blane, Katherine L. Bellis.

¹⁴These authors jointly supervised this work: Sharon J. Peacock, Ewan M. Harrison. ✉ e-mail: fcoll@ibv.csic.es; eh6@sanger.ac.uk

infections such as bacteraemia. Colonisation is an important risk factor for *S. aureus* infection^{3,4}, and it is frequently the strain already colonising an individual that causes the infection^{5,6}.

Whilst a few studies have sought to characterise the adaptive changes that *S. aureus* undergoes during colonisation^{7,8}, our understanding remains incomplete. *S. aureus* persistently colonises ~25% of adults, while others are either never, or only transiently colonised (intermittent carrier)⁹. The genome of *S. aureus* encodes a range of adhesion, immune evasion and antimicrobial resistance factors that, when expressed, allow the bacterium to rapidly adapt to the nasal environment^{10–14}. In addition to changes in gene expression, mutations in the genome of *S. aureus* will also be selected during colonisation if beneficial for survival. This is supported by data from an experimental human challenge model in which persistent carriers preferentially select their own strain, suggesting that *S. aureus* is adapted to the conditions encountered in the colonised individual¹⁵. This likely represents adaption to one or more of: (a) competition with other microbes in the nasal microbiota⁹; (b) nutrient availability in nasal secretions¹⁴; (c) adaption to the host immune response and other physiological variation; (d) spatial variation within the nasal environment (epithelium vs. hair follicles)¹⁶; (e) environmental exposures; and (f) the presence of therapeutic antibiotics and disinfectants (likely more acute in the clinical setting)¹⁷.

S. aureus readily transmits between individuals and strain replacement may take place in persistently colonised individuals¹⁸, meaning that *S. aureus* strains face common selective pressures when adapting to a new individual. Mutations conferring an advantage are therefore expected to be enriched within the same genes, or groups of functionally related genes, across multiple *S. aureus* colonising strains. To test this hypothesis, we analysed the genomes of clonal *S. aureus* isolates sampled from the same individuals to identify evidence of adaptation in recently diverged populations of bacteria. A similar approach has recently been applied to investigate genetic changes that could promote, or be promoted by, invasive infection^{5,8,19}, or associated with persistent or relapsing *S. aureus* bacteraemia^{20–22}.

To differentiate potentially adaptive genetic changes from neutral background mutation, we applied a genome-wide mutation enrichment approach to identify loci in the *S. aureus* genome under parallel and convergent evolution that could represent potential signals of adaptation during colonisation. Our results show that despite limited genetic diversity among colonising isolates of the same individual, multiple genes and pathways show a clear mutational signal of adaptation.

Results

Defining within-host genetic diversity in colonising isolates

To investigate putative adaptive genetic changes in *S. aureus* during colonisation we compiled a genomic dataset from 3497 *S. aureus* colonisation isolates from ten independent studies^{5,23–31}, which included a median of 2 isolates (IQR 2 to 4) from 872 individuals (Supplementary Fig. 1, Supplementary Table 1, Supplementary Data 1). The final dataset, after excluding unrelated (non-clonal) isolates from the same host and poor-quality genomes, consisted of 3060 isolate genomes from 791 individuals, and included 1823 nasal isolates (59.6%), 926 isolates from multi-site screens (30.3%) and 311 isolates from other colonising sites (10.1%). Of these, 1412 isolate genomes (46.1%) were genotypically identified as methicillin-resistant *S. aureus* (MRSA) and the remaining 1648 (53.9%) as methicillin-susceptible (MSSA). The high frequency of MRSA is due to the inclusion of studies that only targeted and screened for MRSA (6 out of 10). Out of the 791 *S. aureus* carriers, 769 (97.2%) were sampled in healthcare settings or nursing homes, and the rest from healthy carriers in other community settings. Most individuals ($n = 701$, 88.6%) were sampled in the UK, followed by Singapore ($n = 62$, 7.8%), Thailand ($n = 17$, 2.1%) and Ireland ($n = 11$, 1.4%). In

terms of timing between samples, 556 individuals (70.3%) had their isolates collected within two months, 126 (15.9%) between two to six months, and 95 (12.0%) between six to twelve months, and 14 (1.8%) more than a year apart.

The genetic diversity between isolates colonising the same individual was low (Supplementary Fig. 2), measured either as the number of single nucleotide polymorphism (SNPs) in the core genome (median 1 SNPs, IQR 0 to 4) or the number of genetic variants (SNPs and small indels) across the whole genome (median 3 variants, IQR 1 to 8). Putative recombination events were detected in the bacterial genomes of 15% of individuals ($n = 117/791$), accounting for 23% of the overall mutation count ($n = 1721/7577$) and were predominantly located ($n = 1367/1721$, 80%) in three prophage recombination hotspots within the reference genome used (NCTC8325)³² (Supplementary Fig. 3). This is consistent with previous studies reporting that most homologous recombination in the core genome of *S. aureus* can be found at or around mobile genetic elements (MGEs)³³.

Genome-wide mutation enrichment analysis identifies evidence of adaptation

To identify loci in the *S. aureus* genome exhibiting evidence of parallel and convergent evolution that could represent potential signals of adaptation during colonisation, we applied a genome-wide mutation enrichment approach (Fig. 1). Using clonal isolates sampled from the same host, we quantified the number of protein-altering mutations (missense, nonsense and frame-shift mutations) within each protein coding sequence (CDS) that arose de novo during *S. aureus* colonisation. We then statistically tested whether this was higher than expected when compared to the rest of the genome after correction for multiple testing.

Out of 2326 CDS tested, only the genes encoding the accessory gene regulator A (*agrA*), the accessory gene regulator C (*agrC*) and the assimilatory nitrite reductase large subunit (*nasD*) showed a statistically significant (p -value < 0.05 after adjusting for multiple testing) enrichment of protein-altering mutations (Fig. 2A). Just below the genome-wide significance level were genes encoding known antibiotic targets: *fusA* encoding the target of fusidic acid³⁴; *dfrA* encoding the target of trimethoprim³⁵; and *pbp2*, which encodes a target of beta-lactams³⁶ (Fig. 2A). The finding of *agr* genes (*agrA* and *agrC*), known to be frequently mutated in *S. aureus* carriers^{37,38}, and that of known antibiotic targets demonstrated the feasibility of our approach in detecting putative adaptive mutations.

To broaden the search for signals of convergent evolution in groups of genes that are functionally related, we counted mutations among all genes belonging to the same transcription unit (operon)³⁹. Out of 1166 operons tested, nine reached statistical significance for an excess of protein-altering mutations (Fig. 2B). These included three operons containing genes that reached statistical significance on their own: U1306 (*nasD*) and U1096 and U1095 (both containing *agrA* and *agrC*); and six additional operons containing CDS that did not reach statistical significance on their own: overlapping operons U605, U606 and U604, all containing the *ileS* gene; the U942 operon harbouring four riboflavin biosynthesis genes (*ribD*, *ribB*, *ribA* and *ribH*); the U254 operon containing genes involved in fatty acid metabolism (*vraA*, *vraB* and *vraC*); and the U331 operon which includes a single hypothetical protein (*SAOUHSC_00704*) (Supplementary Data 2).

At the highest functional level, we aggregated mutations within CDS of the same metabolic process, as defined by well-curated metabolic sub-modules in the *S. aureus* JE2 reference genome⁴⁰. Out of 61 metabolic pathways tested, 11 reached statistical significance for an excess of protein-altering mutations, with 'nitrogen metabolism' and 'riboflavin biosynthesis' pathways being the top two metabolic processes affected (Fig. 2C), demonstrating a clear signal of selection on distinct metabolic processes.

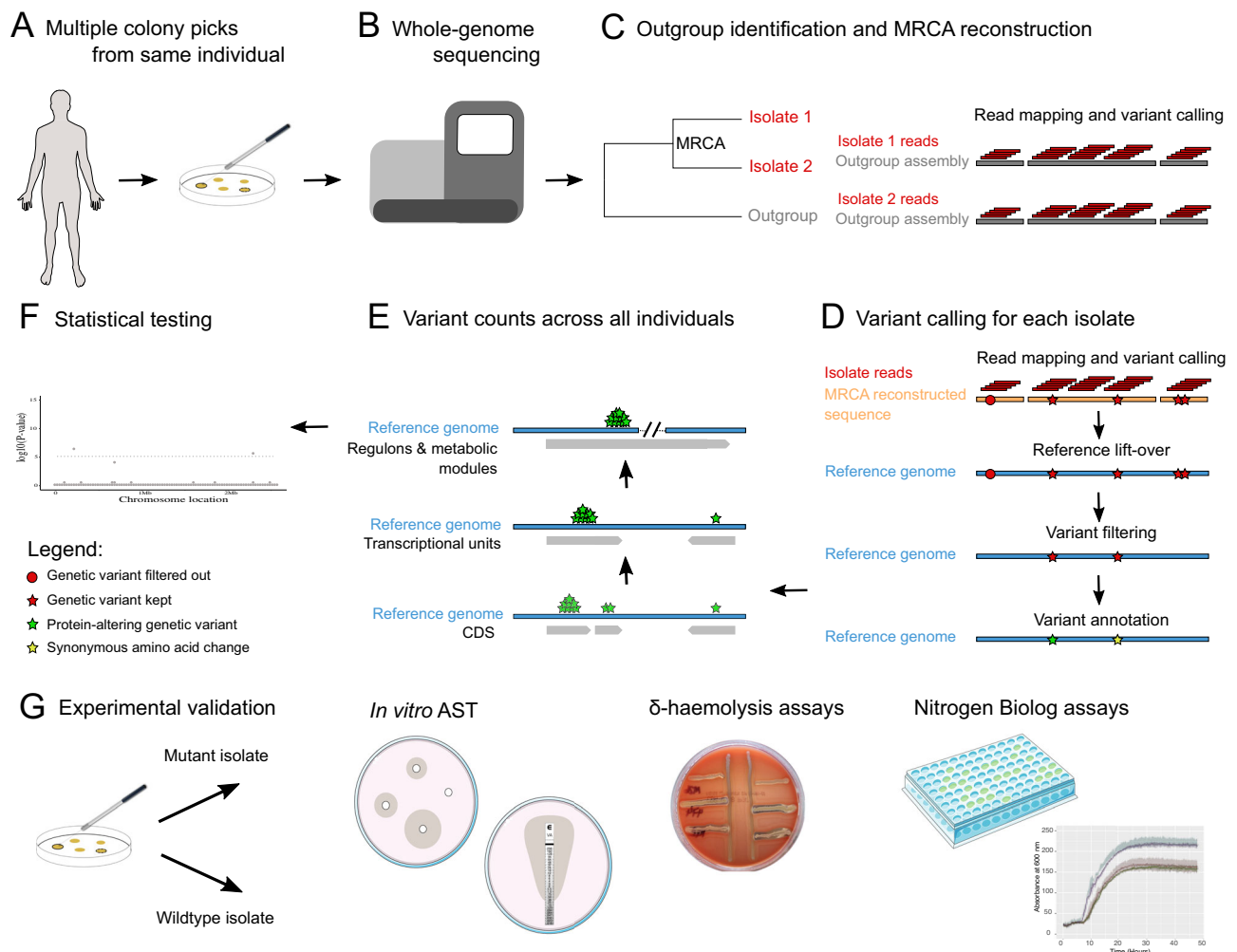


Fig. 1 | Design of genomic analyses to detect potential signals of adaptation.

A *S. aureus* colonies cultured from swabs taken from typical carriage sites of the same individual. (the human body silhouette icon was released into the public domain and obtained from https://commons.wikimedia.org/wiki/File:Human_body_silhouette.svg; the culture plate was created manually; and the inoculation loop obtained from <https://bioicons.com/> under license CC0 1.0). **B** Multiple isolates are whole genome sequenced from the same individual (the sequencing machine icon was created manually). **C** A core-genome phylogeny is used to ensure isolates from the same host are clonal and to identify an appropriate outgroup. Isolate short reads are mapped to the outgroup assembly to call genetic variants. The sequence of the most recent common ancestor (MRCA) of all isolates from the same host is reconstructed. **D** The short reads of each isolate are mapped to the MRCA reconstructed sequence to call variants wherein the reference allele represents the ancestral allele and the alternative allele the evolved one. The coordinates of variants in a complete and well-annotated reference genome (Reference lift-

over) are determined. Variants on repetitive, low-complexity and phage regions are removed as well as those attributable to recombination (Variant filtering). In the last step, the effect of variants on genes is annotated (Variant annotation). **E** The number of protein-altering mutations are counted on protein-coding genes (CDS), transcriptional units (operons) and high-level functional units across all individuals. **F** Each functional unit is tested for an enrichment of protein-altering mutations compared to the rest of the genome. **G** The mutant isolate (with a putative adaptive mutation) and a closely related wildtype isolate obtained from the same individual are tested in vitro for antibiotic susceptibility (AST), delta-haemolytic activity, and growth under a variety of nitrogen sources to validate the phenotypic effect of putative adaptive mutations (the petri dish and multi-well plate icons by Servier <https://smart.servier.com/> were obtained from <https://bioicons.com/> under license CC-BY 3.0; the δ -haemolysis plate and the growth curves were obtained from this work).

Nitrogen metabolic enzymes are enriched for mutations in colonising isolates

Nitrogen metabolism was the metabolic process most enriched by protein-altering mutations in colonisation isolates (Fig. 2C). *nasD* (also named *nirB*) was the third most frequently mutated gene (in a total of 14 individuals), only after *agrA* ($n = 19$) and *agrC* ($n = 20$). *nasD* encodes the large subunit of the assimilatory nitrite reductase, an enzyme responsible for reducing nitrite (NO_2^-) to ammonium, an early step in the fixation of nitrogen from inorganic forms (Fig. 3A). After *nasD*, the gene encoding the urease accessory protein *UreG* (*ureG*), was the second most mutated nitrogen metabolic enzyme (17th hit, Supplementary Data 2). Urease is a nickel-dependent metalloenzyme that

catalyses the hydrolysis of urea into ammonia (NH_3) and carbon dioxide (CO_2)⁴¹.

Because urea is by far the most abundant organic substance in nasal secretions¹⁴, we hypothesised that mutations in *nasD* and *ureG* could represent adaptations to the abundant availability of this nitrogen source. To investigate this, we first tested *nasD* and *ureG* transposon knockouts for their ability to grow under a variety of nitrogen sources. We observed rapid growth with amino acids like glycine, but slower growth with urea and ammonia as the primary nitrogen source, and even slower with nitrate and nitrite (Supplementary Fig. 4, Supplementary Data 3). Compared to the control strain (*comEB* transposon knock-out), the growth of the *nasD* knock-out was compromised in

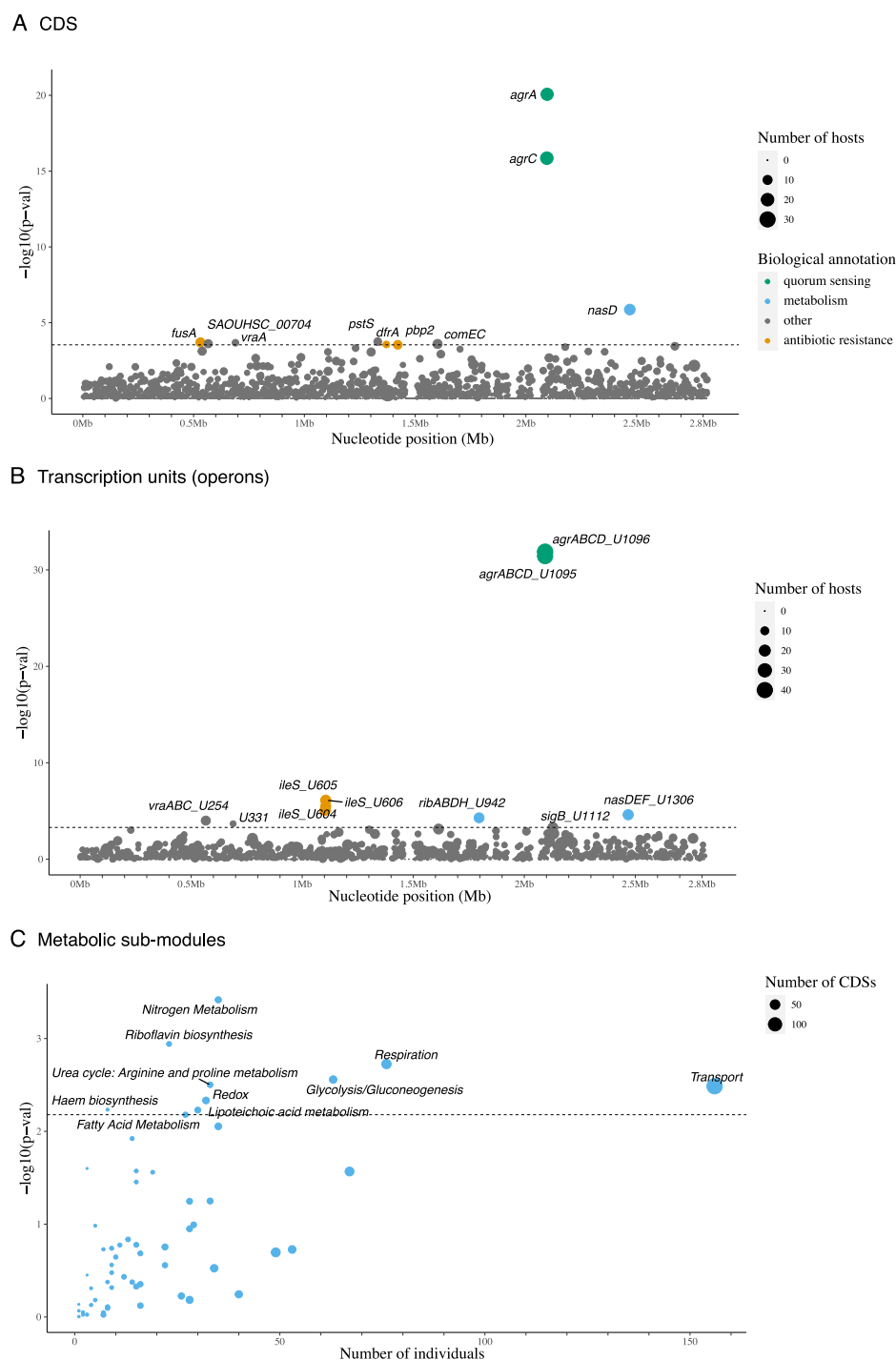


Fig. 2 | Loci enriched for protein-altering mutations in colonising isolates.

A Protein coding sequences (CDS) and **(B)** transcriptional units (operons) enriched for protein-altering mutations in colonising isolates of the same host. Each circle denotes a single locus, whose size is proportional to the number of hosts mutations arose independently from. Loci are placed at the x-axis based on their chromosome coordinates. The y-axis shows the uncorrected p -value resulting from a single-tailed Poisson test comparing the density of mutations per functional unit (CDS, transcription unit and metabolic sub-module) against the expected number of mutations, obtained from multiplying the genome-wide mutation count per bp by the

gene length (see Methods). The dotted horizontal line represents the genome-wide statistical significance threshold. **C** Metabolic sub-modules enriched for protein-altering mutations in colonising isolates. In the x-axis, number of independent acquisitions of protein-altering mutations in different hosts across all protein-coding sequences (CDS) of the same metabolic sub-module. The number of CDS making up metabolic sub-modules is indicated with the size of each circle. In the y-axis, strength of statistic association shown by adjusted p -value. Mutations identified in a collection of 3060 *S. aureus* isolate genomes from 791 individuals (see Supplementary Table 1). Source data are provided as a Source Data file.

multiple nitrogen sources (Fig. 3C), including urea (growth rate 0.32 vs. 0.51, one-way ANOVA p -value < 0.01). Likewise, the growth rate of the *ureG* knock-out was significantly compromised with urea (0.25 vs. 0.51, one-way ANOVA p -value < 0.001 , Supplementary Data 3),

highlighting the critical role of *ureG* in the utilisation of urea as the main nitrogen source for growth.

Next, we tested available colonising isolates with naturally occurring *nasD* mutations (Supplementary Table 2), and their

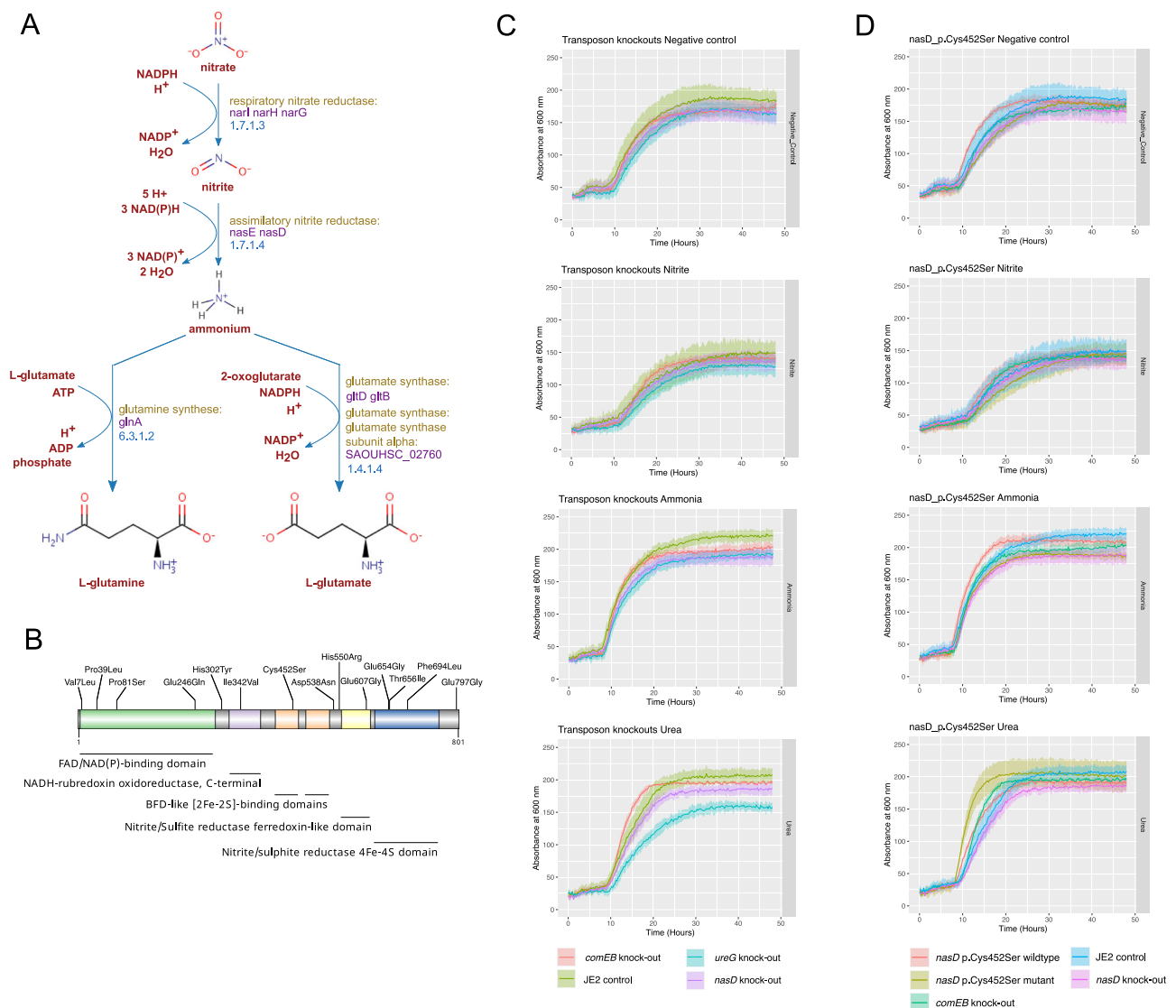


Fig. 3 | Assimilatory nitrite reductase metabolic role, mutations, and nitrogen utilization of mutants. A Role of the assimilatory nitrite reductase enzyme encoded by *nasD* in the nitrate assimilatory pathway of *S. aureus*. Adapted from BioCyc. **B** Location of missense mutations along NasD protein. Pfam protein domains are shown in distinct colours. **C** Growth curves of *S. aureus* *nasD/nirB* knock-out, *ureG* knock-out, *comEB* knock-out (control) and JE (control) strain under the following nitrogen sources: negative control well, nitrite, ammonia, and urea.

D Growth curves of *S. aureus* *nasD/nirB* p.Cys452Ser mutant, wildtype (i.e. quasi-isogenic isolate lacking the *nasD/nirB* mutation from the same host) and controls strains under the same nitrogen sources. Coloured lines represent mean OD600 calculated across three replicates, and shaded coloured regions the standard deviation. See Supplementary Data 5 for all mutations identified between colonising isolates of the same host. Source data are provided as a Source Data file.

corresponding closely related *nasD*-wildtype isolates from the same host ($n = 3$), for growth under the same nitrogen sources. Compared to the wildtype isolate, a Glu246Gln mutant (ST22) showed reduced growth under most nitrogen sources (Supplementary Fig. 5), including in the nitrogen negative control well, though the difference was most pronounced in urea, suggesting the fitness of this mutant was compromised relative to its wildtype. The Thr656Ile mutant (ST22) and wildtype both showed similar growth parameters across nitrogen sources, though the wildtype grew marginally better in urea than the mutant suggesting this mutation would be detrimental to growth in urea. Conversely, the Cys452Ser mutant (ST5) showed a statistically significant improvement in growth compared to its wildtype (in terms of a higher exponential growth rate: 0.64 vs 0.38, p -value < 0.001) in the presence of urea (Fig. 3D), compared to inorganic nitrogen sources. These results point to an adaptive effect of *nasD* Cys452Ser

mutation in the presence of urea. Interestingly, we also observed a strong effect of the strain's genetic background on growth, with ST5 isolates (Supplementary Fig. 5I–L) growing comparably as well as the transposon control strains (ST8), and ST22 isolates growing comparably worse.

Adaptive mutations reveal well-known and unreported anti-biobiotic resistance mutations

Our initial data suggested that the targets of antibiotics from distinct functional classes demonstrate a potential signal of adaptation (Fig. 2A). As such, we investigated whether mutations in these genes reduced susceptibility to their cognate antibiotics (Supplementary Fig. 6) by testing the antibiotic susceptibility of closely related clinical isolates that were mutant and wild-type pairs from the same individual (Fig. 1G). Mutations in *fusA* arose in 10 individuals. Out of the ten

missense variants (Supplementary Table 3), five had the exact amino acid changes previously reported to confer fusidic acid resistance (Val90Ile, Val90Ala, Pro404Leu)⁴² or within the same codon (His457Arg) and were phenotypically resistant to fusidic acid. The other five isolates harbouring *fusA* missense variants were all susceptible to fusidic acid, ruling out an adaptive role of these mutations in fusidic acid resistance.

Five of the eight protein-altering mutations in *ileS* are known (Val588Phe and Val631Phe) or are in a codon (Gly593Ala) known to confer mupirocin resistance⁴² and exhibited elevated MICs compared to the wildtype clonal isolate from the same individual (Supplementary Table 3). We confirmed the role of a not reported frameshift mutation (Ile473fs) in mupirocin resistance (E-test MIC 1024 µg/mL, breakpoint >12 µg/mL) and ruled out the effect of Gly591Ser (E-test MIC 0.5 µg/mL). Out of the five *S. aureus* isolates with missense variants in *dfrA*, three had amino acid changes reported to confer resistance to trimethoprim (His150Arg and two Phe99Tyr)⁴². The available isolate with Phe99Tyr was phenotypically resistant (MIC >= 16 µg/mL), but the isolate carrying His150Arg was not (MIC <= 0.5 µg/mL, zone diameter 27 mm), ruling out the role of this mutation in trimethoprim resistance in this particular strain background.

Missense mutations in *pbp2* were all located within the transglycosylase domain of PBP2 (Supplementary Fig. 6D), which is known to cooperate with PBP2A⁴³ to mediate beta-lactam resistance in MRSA. The three PBP2-mutated strains from available collections²³ were all ST22 (from phylogenetically distinct clades) MRSA (positive for *mecA*/PBP2a), but two were cefoxitin susceptible while retaining benzylpenicillin and oxacillin resistance (Supplementary Table 3). The corresponding PBP2-wildtype isolates from the same individual retain cefoxitin resistance, suggesting these mutations result in cefoxitin susceptibility.

We next investigated two sets of mutations putatively involved in glycopeptide resistance. First, *vraA*, a gene involved in fatty acid metabolism, was the seventh most mutated protein-coding gene ($n = 8$ individuals), and is downregulated in daptomycin tolerant strains⁴⁴. Mutations in other genes involved in cell membrane lipid metabolism (e.g., *mprF/fmtC* or *vraT*, Supplementary Table 4)⁴⁵ are reported to reduce daptomycin susceptibility. Second, *pstS* a gene encoding a phosphate-binding protein, part of the ABC transporter complex PstSACB, was the fourth most frequently mutated protein-coding sequence ($n = 7$ individuals) (Fig. 2A). A point mutation in another phosphate transporter of *S. aureus* (*pitA*) increased daptomycin tolerance⁴⁶. We hypothesised that protein-altering mutations in *vraA* and *pstS* could have a similar effect on daptomycin resistance. We determined daptomycin MICs and tolerance under a sub-inhibitory concentration of daptomycin (0.19 µg/mL) for the available *vraA*-mutated and *pstS*-mutated isolates (Supplementary Table 5), and with *pstS* and *vraA* loss-of-function (LOF) mutations from a larger collection (Supplementary Table 6). These results showed that neither the *pstS* or *vraA* mutations, or LOF mutations led to significant increases in daptomycin MIC, and only the mutant *pstS* p.Gln217* (mean AUC = 10.5, one-way ANOVA p -value < 0.01, Supplementary Data 4, Supplementary Fig. 7) showed increased daptomycin tolerance. The absence of improved growth of mutants relative to controls indicates that the primary driver of *pstS* and *vraA* mutations was not daptomycin tolerance and suggests these mutations could be also metabolic adaptations to fatty acid metabolism.

Agr-inactivating mutations arise frequently in colonising isolates

The genes encoding the sensor kinase AgrC and the response regulator AgrA were, by far, the most frequently mutated genes (Fig. 2A), found in strains colonising 22 and 21 individuals, respectively (including one strain with both an AgrC and AgrA mutation). These genes belong to an

operon encoding the accessory gene regulatory (Agr) system, a two-component quorum-sensing system that senses bacterial cell density and controls the expression of a number of important *S. aureus* virulence factors⁴⁷.

In AgrC, protein-altering mutations were concentrated in the histidine kinase (HK) domain ($n = 16/20$, Fig. 4a), potentially abrogating phosphorylation of AgrA. For AgrA, mutations were enriched in the DNA binding domain ($n = 16/19$, Fig. 4b), likely preventing the binding of phosphorylated AgrA to its cognate DNA binding region. We additionally inspected mutations in the *agr* intergenic region and found that four of the five mutations in this region fall close to the AgrA binding site of Promoter 2 (Fig. 4c). Altogether, these mutations likely abrogate expression of the Agr system by preventing the phosphorylation of AgrA or binding of phosphorylated AgrA to its cognate DNA binding region. To confirm this we tested putative *agr*-defective mutants, and their corresponding *agr*-wildtype isolate from the same host, for delta-haemolytic activity as a proxy for *agr* activity⁴⁸. Given the large number of mutations to test (Supplementary Table 7), we selected 24 isolates from available collections²³ containing a representative mutation (i.e. missense, frameshift, stop gained and inframe indel) at each protein domain or intergenic region. As expected, the selected representative Agr-mutants were negative for delta-haemolytic activity, while their corresponding closely related wildtype isolates retained activity (Fig. 4b). Because *agr* mutants could also have arisen during laboratory passage⁴⁸, we sought to provide evidence of *agr* mutants arising *in host*. The identification of the same mutation in the same strain from independent swabs (i.e., taken at different time points) would point to a single origin of such mutation arising *in host*, as it is less likely for the same mutation to originate multiple independent times in the laboratory. Indeed, we found that all five individuals with multiple Agr-mutated isolates taken from independent swabs (Supplementary Table 7) carried the same Agr mutation.

Mutations that inactivate *agr* have been reported in previous studies, predominantly in *agrC* and *agrA* genes, both in healthy carriers^{4,37,38} and from multiple types of infections⁴⁸, validating our approach to look for signals of adaptation. However, while some studies propose that *agr*-inactivating mutations arise more frequently in infected patients⁵, others report similar frequencies in both infected and uninfected carriers³⁸. To investigate this, we tested whether *agr* mutants were more common in carriers who had staphylococcal infections (206/791) compared to *S. aureus* uninfected carriers (497/791) (See Methods for definitions). We did not find this to be the case (7/206 vs. 10/497, p -value 0.17) after accounting for the number of sequenced isolates, genetic distance, collection, and clonal background as potential confounders (See Methods).

Further putative adaptive mutations in an extended and larger dataset

Our original dataset was compiled in June 2019 (3060 isolates from 791 individuals). To strengthen our initial findings, we searched for newly published studies with multiple colonisation isolates sequenced per individual (up to June 2023), to increase the sample size of the dataset and the chances of detecting additional adaptive variation. We applied the same curation, genomic and QC methodological steps to keep only high-quality and clonal genomes of the same individual from colonisation sources. A total of 4090 additional isolate genomes (63.2% MRSA) obtained from 731 individuals and 15 different studies^{49–63} were included (Supplementary Table 8, Supplementary Data 1). Most individuals were sampled in the USA ($n = 309$, 42.3%), followed by Singapore ($n = 186$, 25.4%), UK ($n = 141$, 19.3%), Japan ($n = 30$, 4.1%), Thailand ($n = 26$, 3.6%) and other countries ($n = 39$, 5.3%). Application of the genome-wide mutation enrichment approach to the combined dataset (7150 isolates from 1593 individuals) revealed even more genes reaching statistical significance for an excess of protein-altering

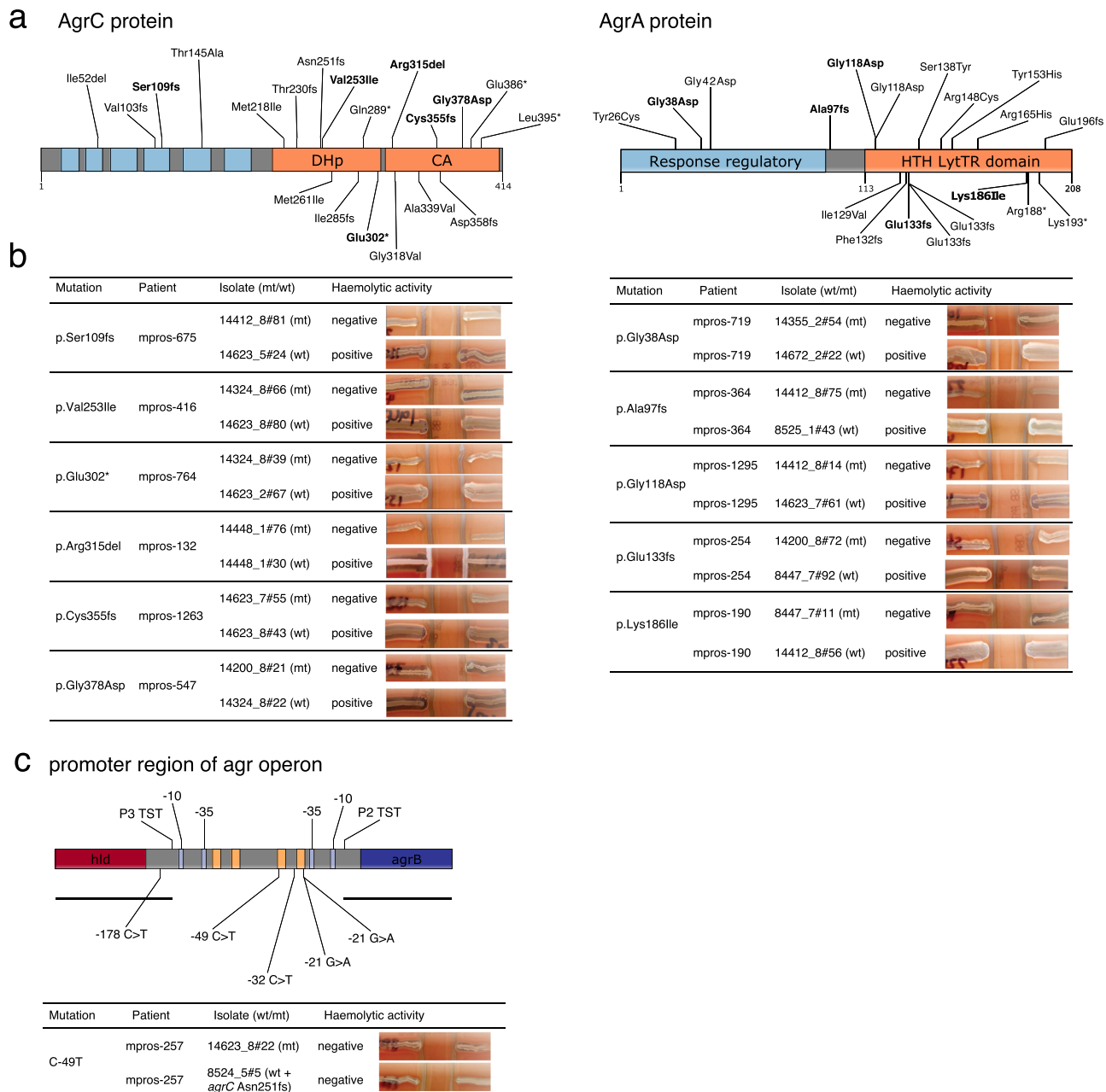


Fig. 4 | Mutations found in the accessory gene regulatory (Agr) system of colonizing strains. a Protein-altering mutations in the protein domains of the sensor kinase AgrC and the response regulator AgrA. The N-terminal sensor domain of AgrC comprises six transmembrane domains (coloured in blue) and is connected to a conserved C-terminal histidine kinase (HK) domain (coloured in orange). The HK domain is made up of two subdomains: the dimerization and histidine phosphotransfer (DHp) subdomain and the catalytic and ATP-binding (CA) subdomain¹¹⁰. AgrA is comprised of a response regulatory domain (coloured in blue) and a DNA binding domain (coloured in orange). Isolates carrying mutations in bold were selected for haemolytic assays from available collections²³ to represent different types of mutations (i.e. missense, frameshift, stop gained and inframe indel) at each protein domain. **b** Haemolytic activities of *S. aureus* isolates on sheep blood agar (SBA) plates used to test the activity of the Agr system. For each mutation, two isolates from the same host were tested, one carrying a selected Agr

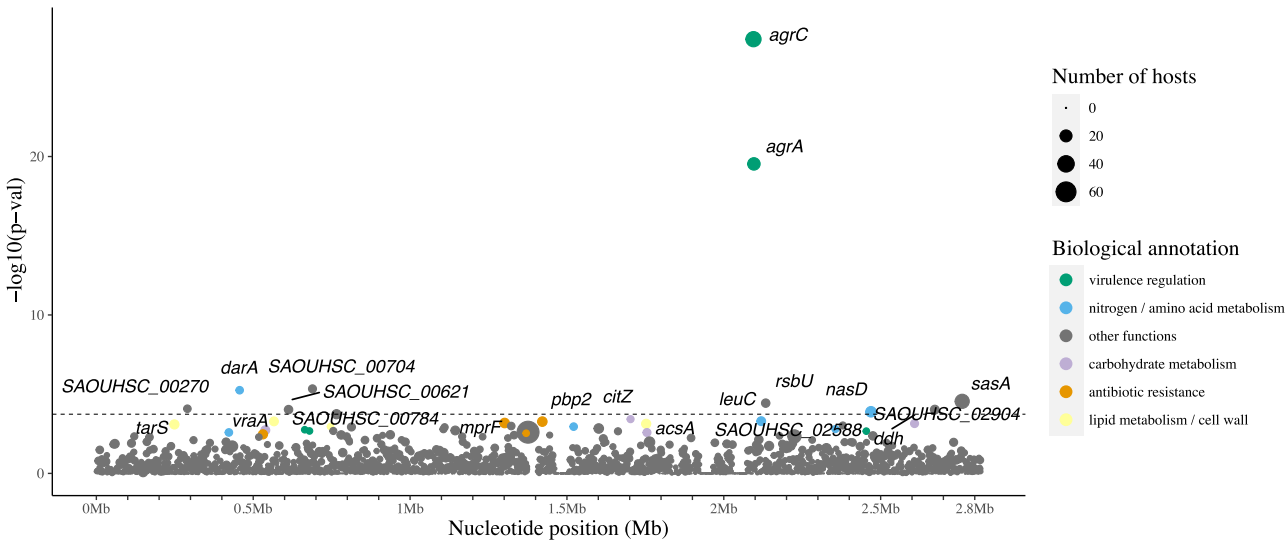
mutation (mutant) and a second isolate being wild type for the Agr system. A positive result is indicated by a widening of haemolysis seen in the region of RN4220. **c** Intergenic region containing agr promoters. The black horizontal lines represent the extent of transcript starting at the promoter 3 transcriptional start site (P3 TST), which encodes for RNAIII, and the transcript starting at promoter 2, which contains the whole *agrBDCA* coding region. Light blue boxes represent -10 and -35 boxes, whereas orange boxes the AgrA binding sites (“AgrA tandem repeats”). The only intergenic mutation carried by an available isolate (C-49T) yielded a negative haemolytic assay, as well as the isolate from the same host lacking this mutation, the latter attributable to a frameshift mutation in AgrC. Mutations identified in a collection of 3,060 *S. aureus* isolate genomes from 791 individuals (see Supplementary Table 1). See Supplementary Data 5 for all mutations identified between colonising isolates of the same host. Source data are provided as a Source Data file.

mutations (Fig. 5, Supplementary Fig. 8), including the ones originally identified (*agrA*, *agrC* and *nasD*) plus an extra eight genes. The latter included *darA/pstA* (which encodes a nitrogen regulatory protein), *sasA* (*S. aureus* surface protein A also known as SraP (serine-rich adhesin for binding to platelets involved in adhesion and invasion)^{64,65},

rsbU (sigmaB regulation protein), and five genes yet to be functionally characterised (SAOUHSC_00704, SAOUHSC_00270, SAOUHSC_00621, SAOUHSC_02904 and SAOUHSC_00784).

The density of protein-altering mutations identified in SraP was highest within the L-lectin domain of the protein (Supplementary

A Top 20 most significant CDS



B Statistically significant CDS (n=11)

locus id	N mutations	gene name	product	p-value
SAOUHSC_02264	35	<i>agrC</i>	accessory gene regulator protein C	1.14×10^{-24}
SAOUHSC_02265	22	<i>agrA</i>	accessory gene regulator protein A	4.23×10^{-17}
SAOUHSC_00704	8	-	conserved hypothetical protein	4.07×10^{-3}
SAOUHSC_00452	7	<i>darA/pstA</i>	nitrogen regulatory protein	4.07×10^{-3}
SAOUHSC_02990	31	<i>sasA/sraP</i>	serine-rich adhesin for platelets	1.63×10^{-2}
SAOUHSC_02301	9	<i>rsbU</i>	sigmaB regulation protein RsbU	1.77×10^{-2}
SAOUHSC_00270	8	-	conserved hypothetical protein	0.03
SAOUHSC_00621	9	-	DMT family transporter	0.03
SAOUHSC_02904	10	-	Ferredoxin-NADP reductase	0.03
SAOUHSC_02684	15	<i>nasD/nirB</i>	assimilatory nitrite reductase, large subunit	0.03
SAOUHSC_00784	11	-	Tetratricopeptide repeat protein	0.04

Fig. 5 | CDS enriched for protein-altering mutations in colonising isolates of the extended dataset. **A** The top 20 most significant CDS are labelled on the plot. Each circle denotes a single locus, whose size is proportional to the number of hosts mutations arose independently from. Loci are placed at the x-axis based on their chromosome coordinates. The y-axis shows their uncorrected *p*-value resulting from a single-tailed Poisson test comparing the density of mutations in CDS against the expected number of mutations, obtained from multiplying the genome-wide mutation count per bp by the CDS length (see Methods). The dotted horizontal line

represents the genome-wide statistical significance threshold. **B** Locus id and annotation of statistically significant CDS (*n* = 11) only. Locus IDs in bold indicate the genes that became statistically significant in the extended dataset. The second column shows the number of mutations originating in different hosts. The *p*-value presented in this table corresponds to the Benjamini-Hochberg corrected *p*-value. Mutations identified in a collection of 7150 *S. aureus* isolate genomes from 1593 individuals (See Supplementary tables 1 and 8). Source data are provided as a Source Data file.

Fig. 9A), a region which has previously been shown to bind to N-acetylneuraminic acid and mediates adherence to human lung epithelial cell lines in vitro⁶⁴. Mapping the position of amino-acid substitutions arising from *sraP* mutations onto the crystal structure of the L-lectin domain revealed that N327 and G329 lie within 4 angstroms of the predicted N-acetylneuraminic acid binding site (Supplementary

Fig. 9B). Amino-acid substitutions at this site have the potential to directly impact ligand binding. Six additional amino-acid substitutions located at a further distance from the ligand binding site could influence N-acetylneuraminic acid binding by altering the protein fold, or could interfere with another as yet unknown function of SraP.

Genes encoding known antibiotic targets (*dfpA*, *fusA* and *pbp2*) remained among the top hits but below the genome-wide significance threshold. Among these was *mprF*, in which point mutations are known to confer daptomycin resistance. These results provide further evidence of the importance of nitrogen metabolism and identifies several uncharacterised genes likely to be critical for colonisation that warrant further experimental investigation.

Discussion

In this study we have provided a comprehensive view of the mutational landscape shaped by selective pressures that *S. aureus* is exposed to during human colonisation. The frequency and type of genomic mutations that arise provide a record of adaptive changes that commensal *S. aureus* underwent in response to evolutionary pressures in the host and provide insights into the biology of *S. aureus* in its primary niche. We compared the genomes of isolates collected from the same host, across a large number of hosts, to detect loci under parallel and convergent evolution⁶⁶.

Our results provided indirect evidence of the ongoing metabolic adaptation of *S. aureus* during colonisation, with the strongest selective pressure being on nitrogen metabolism. We observed that nitrogen metabolic enzymes are often mutated in colonising isolates, specifically genes encoding sub-units of the assimilatory nitrite reductase (*nasD/nirB*) and urease (*ureG*), and a nitrogen regulatory protein (*darA/pstA*) in the extended dataset. Nitrite reduction can also be indicative of growth under anaerobic environments when nitrate (NO₃⁻) and nitrite (NO₂⁻) are used as terminal electron acceptors in place of O₂⁶⁷. Indeed, genes related to dissimilatory nitrate and nitrite reduction are up-regulated under anaerobic conditions⁶⁸, when *nasD/nirB* serves to detoxify the nitrite that accumulates in nitrate-respiring cells⁶⁹. Staphylococcal urease has also been implicated in adaptation to acid environments by ammonia production⁴¹. Therefore, it cannot be ruled out for *nasD/nirB* and *ureG* mutations could represent adaptations to anaerobic and acidic environments, respectively. Adaptions in nitrogen metabolic genes could be a response to distinct nitrogen source availabilities between human hosts, variation in the community structure of the nasal microbiota or between distinct sub-niches within the nasopharynx, therefore requiring metabolic adaptation of *S. aureus* strains when transmitting onto a new host or move to another anatomical niche. All in all, the specific phenotypic and physiological importance of variation in nitrogen metabolic genes is unclear and requires further characterisation.

Riboflavin biosynthesis was the second and eighth most mutated metabolic pathway in the initial and extended datasets, respectively, with *ribD* and *ribA* being the two most mutated genes in this pathway (Supplementary Data 2). Mutations in these genes might result in altered resistance to roseoflavin, an antimetabolite analogue of riboflavin that has antimicrobial properties⁷⁰, but we could not test this hypothesis as we lacked natural mutants of *ribD* and *ribA* from internal collections.

This is the first time that SraP/SasA, a member of the cell wall anchored (CWA) protein family found in Gram-positive bacteria, has been linked to nasal colonisation. *S. aureus* mutants deficient in two unrelated CWA proteins (ClfB and IsdA) were previously shown to be defective in colonising the nasal cavity of rodents due to a reduced ability to adhere to desquamated epithelial cells in the nares^{71–73}. The interaction between SraP and N-acetylneuraminic acid mediates bacterial adherence to human lung epithelial cell lines in vitro⁶⁴ but a role in adherence to desquamated nasal epithelial cells has not been described. Future research will need to focus on determining the functional significance of SraP mutations occurring inside and outside of the L-lectin domain.

The targets of fusidic acid (elongation factor G), trimethoprim (dihydrofolate reductase), mupirocin (isoleucyl-tRNA synthetase) and

beta-lactams (penicillin-binding protein 2) showed a clear signal of adaptation as revealed by the independent emergence of mutations in the *S. aureus* isolates of multiple individuals. This most likely represents examples of directional selection, wherein *S. aureus* adapted to antimicrobial evolutionary pressures in vivo. The ancestral and evolved alleles could be determined using a closely related isolate sampled from a different individual as an outgroup (see Methods). This was supported by the identification of well-known resistance mutations in these genes, and concomitant reduced antibiotic susceptibility in isolates with these mutations, when compared to quasi 'isogenic' wild-type strains isolated from the same host. However, not all mutations detected in AMR loci were likely to be adaptive. This is exemplified by the characterisation of *fusA*, which had five mutations known to be involved in resistance leading to increases in MIC, and five never reported to cause resistance and not affecting fusidic acid susceptibility. It is therefore the excess of adaptive resistance-conferring mutations that increases the statistical significance of *fusA* and that of other AMR genes. We also identified unreported mutations suggesting that the full diversity of resistance mutations to these drugs is yet to be fully understood and warrants further study. The mutations identified in the transglycosylase domain of PBP2, two of which resulted in cefoxitin susceptibility, are consistent with the cooperation of this native PBP with the acquired PBP2A to mediate beta-lactam resistance in MRSA⁴³, and suggests that these might be compensatory mutations to optimise the function of the transglycosylase domain of PBP2. We hypothesised that protein-altering mutations in *vraA* and *pstS* could decrease susceptibility to daptomycin but we could not prove it. *pstS* mutations could be relevant to adaptation to phosphate limitation in colonising sites or in increasing resistance to nitric oxide stress by altering phosphate uptake⁷⁴. The adaptive role of *vraA* and *acsA*, both encoding long chain fatty acid-CoA ligases, needs to be further investigated.

Our results support previous observations that *agr* genetic diversity is selected for during colonisation³⁸. It has been proposed that a balance exists between wild-type and *agr*-defective cells in the population, where the latter, termed as 'cheaters', benefit from the secretions of virulence factors by wild-type cells without having to produce the costly cooperative secretions⁷⁵. However, in the context of colonisation, expression of the *agr* locus results in the down regulation of several surface proteins including cell wall secretory protein (IsaA)⁷⁶ and fibronectin binding protein B (FnBPB)⁷⁷, which are known to be involved in the attachment of *S. aureus* to cells in the nasal epithelium. Given the importance of these proteins to colonisation, it would be beneficial for *S. aureus* populations to maintain sub-populations of cells that are primed for attachment to a new host. Thus, mutations in *agr* most likely represent an example of balancing selection, where the bacterial population as a whole benefits from having both active and defective *agr* systems, as opposed to a case of directional selection. Some caution is required however, as it is known that *agr* mutants arise frequently during the course of infections⁵, and in vitro during post-isolation handling⁴⁸. While we provide evidence of a subset of *agr* mutations arising *in host*, deep metagenomic sequencing of primary samples could be used in future studies to determine if mutations in *agr* are present before culture.

As previously proposed⁸, adaptive antibiotic resistance mutations could result in other phenotypic effects (pleiotropic effects) or be instead the result of immune evasion adaptations. As such, *mprF* mutations reduce susceptibility to cationic antibiotics like daptomycin⁷⁸ but also to host defence cationic peptides produced by neutrophils⁷⁹, representing a potential immune evasion strategy. Similarly, mutations in *tarS*, which encodes the wall teichoic acid (WTA) β -glycosyltransferase and was frequently mutated in the extended dataset, could be adaptive in the context of immune evasion, as it has been shown that *tarS* knockout mutants escape host anti-WTA-IgG mediated opsonophagocytosis⁸⁰. *tarS* variants could also be

influencing susceptibility to staphylococcal phages and beta-lactams antibiotics, as TarS-mediated WTA β -glycosylation is required for susceptibility to podoviruses⁸¹ and methicillin resistance, respectively⁸².

Compared to previous and similar studies reporting on *S. aureus* genetic adaptations, a recent study⁸ also observed enrichments of mutations in *agrA*, *agrC*, *pbp2* and *mprF* between colonisation isolates of the same host. Mutations in genes encoding enzymes of the tricarboxylic acid cycle (TCA), such as *sucA*⁸ and *fumC*⁸³, have also been detected frequently in skin-adapted strains. When investigating infection-specific adaptations, a high frequency of mutations in *agr* genes is often reported⁵, as well as in genes involved in antibiotic resistance (e.g., *rpoB* and *rpsJ*)⁸.

Our study has several limitations. First, the full genetic diversity of *S. aureus* in colonising sites was not captured as only a median of two sequenced colonies were available per individual. Second, we did not investigate changes in the gene content and large genetic rearrangements driven by movement of MGEs between isolate genomes of the same host, as this would require long-read sequencing. We did not explore the role of homologous recombination in adaption. Of note, 80% of mutations attributable to recombination fell within three prophage regions, consistent with previous studies showing that MGEs are hotspots of homologous recombination in the core genome of *S. aureus*³³. These results point to a translocation and movement of Staphylococcal phages during colonization, the adaptive effect of which should be the focus of future studies. Third, we did not have metadata, such as antibiotic usage, or the specific site of colonisation (e.g., multi-site screens) for 30.3% of isolates. Fourth, most isolates came from studies of *S. aureus* colonisation in hospital patients with a bias towards MRSA strains, which may have incorporated a bias towards mutations selected by antibiotics or other therapies. Fifth, some of the mutational adaptations identified here could well represent adaptations of *S. aureus* strains to longer or extended periods of colonization, as those expected in persistent carriers, as these strains are more likely to be isolated over multiple sampling points and hence to be included in this study. Finally, generating and testing clean isogenic mutants would have provided stronger and more conclusive evidence about the role of individual mutations on the tested phenotypes. By increasing the overall sample size from ~3000 to ~7000 genomes we identified new genes significantly enriched for mutations including five currently uncharacterised genes and *sasA/sraP* which has not been previously been reported to be involved in colonisation, though it is known to mediate attachment to human cells⁶⁴. This suggests that studies using even larger sample sizes have the potential to identify further new signatures of adaption.

Future work focused on pre-defined patient groups (healthy colonised individuals), narrowly defined infection types^{53,84} with larger sample sizes and availability of host metadata will improve the identification of bacterial adaptive changes that promote survival in specific host niches and in vivo conditions; as well as defining strain/lineage-specific adaptations. Larger samples sizes will also allow us to determine which genes are essential for growth in different conditions, as shown by genes that are rarely inactivated.

While adaptation of clinical *S. aureus* strains during infection has been the focus of multiple recent studies^{5,22,85–87}, to our knowledge, this is the first comprehensive study to investigate adaptation of *S. aureus* populations during human colonisation. Our analysis has identified numerous metabolic pathways and genes likely critical to *S. aureus* colonisation that have not been previously reported and demonstrated the functional impact of these mutations. Our data now warrant detailed experimental investigations to further elucidate *S. aureus* biology during colonisation. Finally, it is likely that our approach can be applied to other bacterial species with similar success.

Methods

Strain collections and data curation

We identified available collections of *S. aureus* genomes with multiple carriage isolates sequenced from the same human individual (Supplementary Data 1)^{5,23–31}. The NCBI Short Read Archive (SRA) was systematically queried on June 2019 to identify BioProjects that met the following criteria (Fig. 1): contained *S. aureus* genomic sequences, could be linked to a publication, included genomes of clinical isolates, clinical sources were known, multiple colonising isolates per host were sequenced, and host IDs were available. Only isolates from colonisation specimens were kept, that is, from multi-site screens^{23,27,29,30} and typical colonising anatomical sites (nose^{5,28,31}, armpit, groin, perineum and throat)^{24–26}. Colonised hosts were classified as symptomatic or asymptomatic carriers based on whether they had a *S. aureus* infection or not, respectively. In studies where clinical specimens were systematically collected from recruited cases^{24,26,30}, individuals were labelled as asymptomatic carriers unless having a clinical specimen collected. In other studies, carriers were all explicitly referred to as infected⁵ or uninfected^{30,31}. In one study, only the nasal carriage controls were kept, as were thus labelled as uninfected. In the rest of studies, no information was available to determine their *S. aureus* infection status^{25,27,29}, and were thus labelled as ‘unknown’.

Genomic analyses applied to all isolates

The Illumina short reads of all *S. aureus* genomes were validated using *fastqcheck* v1.1 (<https://github.com/VertebrateResequencing/fastqcheck>) and de novo assembled using Spades v3.12.0⁸⁸ to create draft assemblies. These were then corrected using the bacterial assembly and improvement pipeline⁸⁹ to generate improved assemblies. All assemblies were evaluated using QUAST v5.0.1⁹⁰ and reads mapped back to de novo assemblies to investigate polymorphisms (indicative of mixed cultures) using Bowtie2 v1.2.2⁹¹. Kraken v2.1.2 was used to determine the proportion of *S. aureus* reads in raw fastq files. Low-quality genomes were excluded from further analysis applying the following thresholds: *S. aureus* reads <80%, N50 <10000, contigs smaller than 1 kb contributing to >15% of the total assembly length, total assembly length outside of the median \pm one standard deviation, or >1500 polymorphic sites.

Methicillin genotypic resistance was determined from improved assemblies using AMRFinderPlus⁹² v3.11.11 (AMRFinder database v2023-08-08.2). Sequence types (STs) were derived from improved assemblies by extracting all seven *S. aureus* multi-locus sequence type (MLST) loci and comparing them to the PubMLST database (www.pubmlst.org)⁹³. Clonal complexes (CCs) were derived from these allelic profiles, allowing up to two allele mismatches from the reference ST. The short reads of each isolate were mapped to the same reference genomes (CC22 HO 5096 0412 strain, accession number HE681097) using *SMALT* v0.7.6 (<http://www.sanger.ac.uk/resources/software/smalt/>), whole-genome alignments were created by calling nucleotide alleles along the reference genome using *SAMtools* and *bcftools* v0.1.19⁹⁴. We kept the portion of the reference genome corresponding to the *S. aureus* core genome in whole genome alignments to calculate core-genome pairwise SNP distances using *pairsn* v0.0.1 (<https://github.com/gtonkinhill/pairsn>). The core genome of *S. aureus*⁹⁵ was derived from an independent, genetically and geographically diverse collection of 800 *S. aureus* isolates genomes from multiple host species⁹⁶ using *Roary*⁹⁷ v3.11.1 with default settings. Core-genome alignments were used to construct a maximum likelihood phylogeny for each clonal complex using *RAxML* v8.2.8⁹⁸ with 100 bootstraps.

Genomic analyses applied to isolates of the same host

To avoid comparing the genomes of divergent strains from the same individual, only clonal isolates were kept for further analyses. Clonality was ruled out if isolates belonged to different clonal complexes or to

the same clonal complex separated by more than 100 SNPs. Clonality was ruled in if isolates differed by less than the maximum within-host diversify previously reported (40 SNPs)⁹⁹. Clonality was investigated for the remaining isolates pairs (differing between 40 to 100 SNPs) by making sure they all clustered within the same monophyletic clade in the phylogenetic tree.

The nucleotide sequence of the most recent common ancestor (MRCA) of all isolates of the same host was reconstructed first. To do this, we used the maximum likelihood phylogenies to identify, for each individual, the most closely related isolate sampled from a different individual that could be used as an outgroup. We used the de novo assembly of this outgroup isolate as a reference genome to map the short reads of each isolate, call genetic variants (SNPs and small indels) using *Snippy* v4.3.3 (<https://github.com/tseemann/snippy>), and build within-host phylogenies using *RAXML* phylogeny and rooted on the outgroup. The ancestral allele of all genetic variants at the internal node representing the MRCA of all isolates of the same host was reconstructed using *PastML* v1.9.20¹⁰⁰. This reconstructed ancestral sequence was used as the ultimate reference genome to call genetic variants (SNPs and small indels). This pipeline was implemented in four python scripts (`identify_host_ancestral_isolate.step1.py` to `identify_host_ancestral_isolate.step4.py`) available at <https://github.com/francescoll/staph-adaptive-mutations>.

As variants were called in a different reference genome for each individual's *S. aureus* strain, they had to be transferred to the same reference genome to allow comparison and annotation of genetic variants from all individuals' strains. We modified an already published script (*insert_variants.pl*)⁶⁶ to find the genome coordinates of variants in the NCTC8325 (GenBank accession number NC_007795.1) and JE2 (NZ_CP020619.1) reference genomes. This script takes a 200-bp window around each variant in one reference (assembly) and finds the coordinates of this sequence in a new reference using *BLASTN*¹⁰¹ and *bcftools* v1.9⁹⁴. Because of this requirement, variants at the edge of contigs (200 bp) were filtered out. The script was modified to keep the single best blast hit of each variant, meaning that variants with window sequences mapping to repetitive regions of the reference genome were removed. Variants in repetitive regions, detected by running *Blastn* v2.8.1+ on the reference genome against itself, and variants in regions of low complexity, as detected by *dustmasker* v1.0.0¹⁰² using default settings, were also filtered out. The final set of high-quality variants were annotated using *SnpEff* v4.3¹⁰³ in both the NCTC8325 and JE2 reference genomes. *Breseq* v0.39.0¹⁰⁴ was run to identify mutations and large indels between the pairs of related isolates from the same individual that were tested in vitro (Supplementary Data 5).

Genome-wide mutation enrichment analysis

To scan for potential adaptive genetic changes recurrent across multiple individuals, we counted the number of functional mutations (i.e., those annotated as having HIGH or MODERATE annotation impact by *SnpEff*) in well-annotated functional loci across all individuals. Before that, putative recombination events, identified as variants clustered within a 1000-bp window in isolate genomes of the same host, were filtered out to avoid inflating mutation counts. When more than two isolates from the same host were available, we made sure the same mutations, identified in multiple case-control pairs of the same host, were counted only once.

We aggregated protein-altering mutations within different functional units. At the lowest level, we counted mutations within each protein coding sequence (CDS). To increase the power of detecting adaptive mutations in groups of genes that are functionally related, we aggregated mutations within transcription units (operons). The coordinates of transcription start and termination sites in the NCTC8325 reference genome were extracted from a study that comprehensively characterised the transcriptional response of *S. aureus* across a wide range of experimental conditions^{39,105}. To our knowledge, this is the

best characterised reconstruction of transcriptional units in *S. aureus*. At the highest functional level, we aggregated mutations within CDS of the same metabolic process, as defined by well-curated metabolic sub-modules in the JE2 reference genome⁴⁰.

We tested each functional unit (CDS, transcription unit and metabolic sub-module) for an excess of protein-altering (functional) mutations compared to the rest of the genome, considering the length of CDS, or cumulative length of CDS if testing high-order functional units involving multiple CDS. To do this, we performed a single-tailed Poisson test using the genome-wide mutation count per bp multiplied by the gene length as the expected number of mutations as previously implemented⁶⁶. Annotated features shorter than 300 bp long were not tested. *P* values were corrected for multiple testing using a Benjamini & Hochberg correction using the total number of functional units in the genome as the number of tests. We chose a significance level of 0.05 and reported hits with an adjusted *P* value below this value, unless otherwise stated.

Other statistical analyses

We tested whether the presence of *agr* mutants, defined as isolates with protein-altering mutations in either *agrA* or *agrC*, was affected by hosts having an *S. aureus* infection (infection status). We fitted a binomial generalized linear model (GLM) using the presence of *agr* mutants as the binary response variable and *S. aureus* infection status as a binary predictor variable. We additionally included the number of sequenced isolates per host, genetic distance of these (expressed as the number of core-genome SNPs), collection and clonal background (clonal complex) as covariates to control for the effect of these potential confounders. This was implemented using the “glm” function (family binomial) in the base package within the statistical programming environment R version 3.4.1¹⁰⁶. The only predictors that increased the odds of detecting *agr* mutants were the number of sequenced isolates per host (odds ratio 1.20, 1.10 to 1.34 95% confidence interval, *p*-value < 0.001) and their genetic distance (odds ratio 1.06, 95% confidence interval 1.01 to 1.11, *p*-value < 0.05).

In vitro antibiotic susceptibility testing

Isolates from frozen stocks were grown overnight on Columbia blood agar (CBA, Oxoid, UK) at 37 °C. Fusidic acid or trimethoprim susceptibility testing was performed using disc (Oxoid, UK) diffusion as per EUCAST recommendations¹⁰⁷. Minimum inhibitory concentration (MIC) testing was performed for daptomycin, vancomycin and mupirocin. A loopful of the isolate added to phosphate buffered saline (PBS), adjusted to 0.5 McFarland, then a thin layer spread evenly on a Muller Hinton agar plate (Oxoid, UK). An antimicrobial gradient strip (Biomérieux, France) was carefully placed, then the plate incubated overnight at 37 °C. The MIC was interpreted as the value on the strip above the point where growth stops.

Biolog experiments

Isolates from frozen stocks were plated on to Lysogeny broth (LB) agar and grown overnight at 37 °C. For the transposon knock-out strains, obtained from Nebraska transposon mutant library, the plates included 5 µg/ml erythromycin. A damp swab was used to take sufficient colonies to create three 81% (+/- 2%) turbidity solutions for each strain in 20 ml PBS. 1.28 ml of each turbid solution was added to 14.83 ml 1.2x IFOa (77268, Biolog), redox dye H (74228, Biolog), and PM3 Gram Positive Additive (made as described by the Biolog protocol). Each well of a PM3 plate (12121, Biolog) was inoculated with 150 µl of this solution. The inoculated plates were run on the Omnilog (Biolog) for 48 h at 37 °C. Readings were taken every 15 min.

Delta-haemolysis experiments

The δ -haemolysis assay was performed as previously described⁴⁸. A thin streak of *S. aureus* strain RN4220 was placed down the centre of a

sheep blood agar plate. A thin streak of the test strain was placed horizontally up to, but not touching, RN4220. Test strains were tested in duplicate. Plates were incubated at 37 °C for 18 h, then at 4 °C for 6 h. Enhanced lysis by the test strain in the area near to RN4220 was an indicator of δ -haemolysis production.

Growth curves

Test isolates were grown overnight at 37 °C in tryptic soy broth (TSB) with 5 μ l/ml erythromycin (transposons) or TSB alone (non-transposons). The overnight cultures were then diluted 1/1000 in minimal media (1 \times M9 salts, 2 mM MgSO₄, 0.1 mM CaCl₂, 1% glucose, 1% caseaminoacids, 1 mM thiamine hydrochloride and 0.05 mM nicotinamide) with 0.095 μ g/ml daptomycin. 300 μ l was added to a 96-well plate, then placed on a FluoStar Omega (BMG Labtech, Germany) for 24 h incubation with shaking. Optical density measurement at OD₆₀₀ was taken every 30 min, and standard curves produced. Each isolate was tested in biological and measurement triplicate.

The R scripts used to process raw growth data, plot growth curves, fit growth curves and compare growth parameters are available on GitHub (https://github.com/francescoll/staph-adaptive-mutations/tree/main/growth_curves). Raw growth data (i.e. absorbance values at different time points) was processed with script `prepare_growth_curves_data.R`. Mean OD₆₀₀ values and 95% confidence limits around the mean were plotted using `ggplot2`¹⁰⁸ functions in script `plot_growth_curves.R`. Growth curves were fitted with `Growthcurver`¹⁰⁹ and growth parameters (growth rate and area under the curve) extracted using script `fit_and_plot_growth_curves.R`. Due to the prolonged lag phase in curves obtained under daptomycin exposure, these curves were fitted after 7 h. We fitted logistic curves to each replicate ($n = 9$) using `Growthcurver` package in R and extracted the growth rate and area under the logistic curve from fitted curves. These growth parameters were compared between isolates/strains (e.g. mutant vs. wildtype) using a one-way ANOVA to determine whether there were any statistically significant differences between the means (across replicates) of growth parameters between isolates (script: `compare_growth_parameters.R`).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The whole genome sequences of the initial and additional isolate collections used in this study are available on European Nucleotide Archive (ENA) under the accessions listed in Supplementary Data 1, which also includes isolate metadata. The initial dataset includes isolate genomes obtained from BioProject accessions PRJEB11177, PRJEB20148, PRJEB2076, PRJEB2655, PRJEB2756, PRJEB3174, PRJEB4141, PRJEB7654, PRJEB9390, PRJNA270998, PRJNA324190, and PRJNA369475; and the additional datasets from PRJDB5246, PRJEB2076, PRJEB28206, PRJEB33854, PRJEB40888, PRJEB4140, PRJEB43023, PRJEB9390, PRJNA530184, PRJNA587530, PRJNA590514, PRJNA595570, PRJNA638400, PRJNA685142, PRJNA715375, PRJNA715649, PRJNA816913, and PRJNA918392. The full list of protein-coding regions, transcriptional units and metabolic processes enriched by protein-altering mutations can be found in Supplementary Data 2. Supplementary Data 3 and 4 include the data of bacterial growth curves. Source data are provided with this paper.

Code availability

All scripts necessary to run the described analyses are available on GitHub (<https://github.com/francescoll/staph-adaptive-mutations>) and Zenodo (<https://zenodo.org/records/14168960>, <https://doi.org/10.5281/zenodo.14168960>)

References

- Kluytmans, J., van Belkum, A. & Verbrugh, H. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin. Microbiol. Rev.* **10**, 505–520 (1997).
- Piewngam, P. & Otto, M. *Staphylococcus aureus* colonisation and strategies for decolonisation. *Lancet Microbe* **5**, e606–e618 (2024).
- Bode, L. G. M. et al. Preventing Surgical-Site Infections in Nasal Carriers of *Staphylococcus aureus*. *N. Engl. J. Med.* **362**, 9–17 (2010).
- von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. Nasal Carriage as a Source of *Staphylococcus aureus* Bacteremia. *N. Engl. J. Med.* **344**, 11–16 (2001).
- Young, B. C. et al. Severe infections emerge from commensal bacteria by adaptive evolution. *eLife* **6**, 1–25 (2017).
- Benoit, J. B., Frank, D. N. & Bessesen, M. T. Genomic evolution of *Staphylococcus aureus* isolates colonizing the nares and progressing to bacteremia. *PLoS ONE* **13**, 1–18 (2018).
- Goyal, M. et al. Genomic Evolution of *Staphylococcus aureus* During Artificial and Natural Colonization of the Human Nose. *Front. Microbiol.* **10**, 1–10 (2019).
- Giulieri, S. G. et al. Niche-specific genome degradation and convergent evolution shaping *Staphylococcus aureus* adaptation during severe infections. *eLife* **11**, 1–33 (2022).
- Krismer, B., Weidenmaier, C., Zipperer, A. & Peschel, A. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. *Nat. Rev. Microbiol.* **15**, 675–687 (2017).
- Thammavongsa, V., Kim, H. K., Missiakas, D. & Schneewind, O. *Staphylococcal* manipulation of host immune responses. *Nat. Rev. Microbiol.* **13**, 529–543 (2015).
- Burian, M. et al. Temporal Expression of Adhesion Factors and Activity of Global Regulators during Establishment of *Staphylococcus aureus* Nasal Colonization. *J. Infect. Dis.* **201**, 1414–1421 (2010).
- Burian, M. et al. Expression of staphylococcal superantigens during nasal colonization is not sufficient to induce a systemic neutralizing antibody response in humans. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 251–256 (2012).
- Nitzan, M. et al. A defense-offense multi-layered regulatory switch in a pathogenic bacterium. *Nucleic Acids Res.* **43**, 1357–1369 (2015).
- Krismer, B. et al. Nutrient Limitation Governs *Staphylococcus aureus* Metabolism and Niche Adaptation in the Human Nose. *PLoS Pathog.* **10**, e1003862 (2014).
- Belkum, et al. Reclassification of *Staphylococcus aureus* Nasal Carriage Types. *J. Infect. Dis.* **199**, 1820–1826 (2009).
- Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* **12**, 49–62 (2014).
- Fitzgerald, J. R. Evolution of *Staphylococcus aureus* during human colonization and infection. *Infect., Genet. Evolution* **21**, 542–547 (2014).
- Votintseva, A. A. et al. Multiple-Strain Colonization in Nasal Carriers of *Staphylococcus aureus*. *J. Clin. Microbiol.* **52**, 1192–1200 (2014).
- Räz, A. K. et al. Limited Adaptation of *Staphylococcus aureus* during Transition from Colonization to Invasive Infection. *Microbiol. Spectr.* **11**, e0259021 (2023).
- Das, S. et al. Natural mutations in a *Staphylococcus aureus* virulence regulator attenuate cytotoxicity but permit bacteremia and abscess formation. *Proc. Natl Acad. Sci.* **113**, E3101–E3110 (2016).
- Richards, R. L. et al. Persistent *Staphylococcus aureus* Isolates from Two Independent Cases of Bacteremia Display Increased Bacterial Fitness and Novel Immune Evasion Phenotypes. *Infect. Immun.* **83**, 3311–3324 (2015).

22. Giulieri, S. G. et al. Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent *Staphylococcus aureus* bacteraemia. *Genome Med.* **10**, 65 (2018).
23. Coll, F. et al. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci. Transl. Med.* **9**, eaak9745 (2017).
24. Price, J. R. et al. Transmission of *Staphylococcus aureus* between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. *Lancet Infect. Dis.* **17**, 207–214 (2017).
25. Chow, A. et al. MRSA Transmission Dynamics Among Inter-connected Acute, Intermediate-Term, and Long-Term Healthcare Facilities in Singapore. *Clin. Infect. Dis.* **64**, S76–S81 (2017).
26. Price, J. R. et al. Whole-Genome Sequencing Shows That Patient-to-Patient Transmission Rarely Accounts for Acquisition of *Staphylococcus aureus* in an Intensive Care Unit. *Clin. Infect. Dis.* **58**, 609–618 (2014).
27. Tong, S. Y. C. et al. Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res.* **25**, 111–118 (2015).
28. Harkins, C. P. et al. The Microevolution and Epidemiology of *Staphylococcus aureus* Colonization during Atopic Eczema Disease Flare. *J. Investigative Dermatol.* **138**, 336–343 (2018).
29. Tosas Auguet, O. et al. Evidence for Community Transmission of Community-Associated but Not Health-Care-Associated Methicillin-Resistant *Staphylococcus aureus* Strains Linked to Social and Material Deprivation: Spatial Analysis of Cross-sectional Data. *PLOS Med.* **13**, e1001944 (2016).
30. Harrison, E. M. et al. Transmission of methicillin-resistant *Staphylococcus aureus* in long-term care facilities and their related healthcare networks. *Genome Med.* **8**, 102 (2016).
31. Paterson, G. K. et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat. Commun.* **6**, 6560 (2015).
32. Gillaspay, A. F. et al. The *Staphylococcus aureus* NCTC 8325 Genome. in *Gram-Positive Pathogens* 381–412 (ASM Press, 2014). <https://doi.org/10.1128/9781555816513.ch32>.
33. Everitt, R. G. et al. Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. *Nat. Commun.* **5**, 3956 (2014).
34. O'Neill, A. J., McLaws, F., Kahlmeter, G., Henriksen, A. S. & Chopra, I. Genetic Basis of Resistance to Fusidic Acid in *Staphylococci*. *Antimicrobial Agents Chemother.* **51**, 1737–1740 (2007).
35. Vickers, A. A., Potter, N. J., Fishwick, C. W. G., Chopra, I. & O'Neill, A. J. Analysis of mutational resistance to trimethoprim in *Staphylococcus aureus* by genetic and structural modelling techniques. *J. Antimicrobial Chemother.* **63**, 1112–1117 (2009).
36. Long, S. W. et al. PBP2a mutations causing high-level ceftaroline resistance in clinical methicillin-resistant *Staphylococcus* isolates. *Antimicrobial Agents Chemother.* **58**, 6668–6674 (2014).
37. Shopsis, B. et al. Prevalence of *agr* Dysfunction among Colonizing *Staphylococcus aureus* Strains. *J. Infect. Dis.* **198**, 1171–1174 (2008).
38. Smyth, D. S. et al. Nasal carriage as a source of *agr*-defective *staphylococcus aureus* bacteremia. *J. Infect. Dis.* **206**, 1168–1177 (2012).
39. Mäder, U. et al. *Staphylococcus aureus* Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. *PLOS Genet.* **12**, e1005962 (2016).
40. Seif, Y. et al. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLoS Computational Biol.* **15**, e1006644 (2019).
41. Zhou, C. et al. Urease is an essential component of the acid response network of *Staphylococcus aureus* and is required for a persistent murine kidney infection. *PLOS Pathog.* **15**, e1007538 (2019).
42. Kumar, N. et al. Evaluation of a fully automated bioinformatics tool to predict antibiotic resistance from MRSA genomes. *J. Antimicrobial Chemother.* **75**, 1117–1122 (2020).
43. Pinho, M. G., De Lencastre, H. & Tomasz, A. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant *staphylococci*. *Proc. Natl Acad. Sci. (USA)* **98**, 10886–10891 (2001).
44. Camargo, I. L. B. D. C., Neoh, H.-M., Cui, L. & Hiramatsu, K. Serial Daptomycin Selection Generates Daptomycin-Nonsusceptible *Staphylococcus aureus* Strains with a Heterogeneous Vancomycin-Intermediate Phenotype. *Antimicrobial Agents Chemother.* **52**, 4289–4299 (2008).
45. Hines, K. M. et al. Occurrence of cross-resistance and β -lactam seesaw effect in glycopeptide-, lipopeptide- and lipoglycopeptide-resistant MRSA correlates with membrane phosphatidylglycerol levels. *J. Antimicrobial Chemother.* **75**, 1182–1186 (2020).
46. Mechler, L. et al. A Novel Point Mutation Promotes Growth Phase-Dependent Daptomycin Tolerance in *Staphylococcus aureus*. *Antimicrobial Agents Chemother.* **59**, 5366–5376 (2015).
47. Painter, K. L., Krishna, A., Wigneshweraraj, S. & Edwards, A. M. What role does the quorum-sensing accessory gene regulator system play during *Staphylococcus aureus* bacteremia? *Trends Microbiol.* **22**, 676–685 (2014).
48. Traber, K. E. et al. *agr* function in clinical *Staphylococcus aureus* isolates. *Microbiology* **154**, 2265–2274 (2008).
49. Hall, M. D. et al. Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *eLife* **8**, 124–131 (2019).
50. Chow, A. et al. Comparative epidemiology and factors associated with major healthcare-associated methicillin-resistant *Staphylococcus aureus* clones among interconnected acute-, intermediate- and long-term healthcare facilities in Singapore. *Clin. Microbiol. Infect.* **27**, 785.e9–785.e16 (2021).
51. Popovich, K. J. et al. Frequent Methicillin-Resistant *Staphylococcus aureus* Introductions Into an Inner-city Jail: Indications of Community Transmission Networks. *Clin. Infect. Dis.* **71**, 323–331 (2020).
52. Sansom, S. E. et al. Genomic Update of Phenotypic Prediction Rule for Methicillin-Resistant *Staphylococcus aureus* (MRSA) USA300 Discloses Jail Transmission Networks with Increased Resistance. *Microbiol. Spectr.* **9**, e0037621 (2021).
53. Key, F. M. et al. On-person adaptive evolution of *Staphylococcus aureus* during treatment for atopic dermatitis. *Cell Host Microbe* **31**, 593–603.e7 (2023).
54. Annavajhala, M. K. et al. Genomic and Epidemiological Features of Two Dominant Methicillin-Susceptible *Staphylococcus aureus* Clones from a Neonatal Intensive Care Unit Surveillance Effort. *mSphere* **7**, e0040922 (2022).
55. Raghuram, V. et al. Comparison of genomic diversity between single and pooled *Staphylococcus aureus* colonies isolated from human colonization cultures. *Microb. Genomics* **9**, 001111 (2023).
56. van Tonder, A. J. et al. Colonization and transmission of *Staphylococcus aureus* in schools: a citizen science project. *Microb. Genomics* **9**, mgen000993 (2023).
57. Nakamura, Y. et al. *Staphylococcus Agr* virulence is critical for epidermal colonization and associates with atopic dermatitis development. *Sci. Transl. Med.* **12**, eaay4068 (2020).
58. Effelsberg, N. et al. Genotypic Characterization of Livestock-Associated Methicillin-Resistant *Staphylococcus aureus* Isolates of Clonal Complex 398 in Pigsty Visitors: Transient Carriage or Persistence? *J. Clin. Microbiol.* **58**, e01276–19 (2019).

59. Holt, D. C. et al. Longitudinal whole-genome based comparison of carriage and infection associated *Staphylococcus aureus* in northern Australian dialysis clinics. *PLOS ONE* **16**, e0245790 (2021).
60. Ingham, A. C. et al. Dynamics of the Human Nasal Microbiota and *Staphylococcus aureus* CC398 Carriage in Pig Truck Drivers across One Workweek. *Appl. Environ. Microbiol.* **87**, e0122521 (2021).
61. Lydecker, A. D. et al. Targeted gown and glove use to prevent *Staphylococcus aureus* acquisition in community-based nursing homes: A pilot study. *Infect. Control Hospital Epidemiol.* **42**, 448–454 (2021).
62. Lee, R. S. et al. 561. Genomic Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Two Cohorts of High-Risk Military Trainees. *Open Forum Infect. Dis.* **6**, S266–S266 (2019).
63. Popovich, K. J. et al. MRSA Transmission in Intensive Care Units: Genomic Analysis of Patients, Their Environments, and Healthcare Workers. *Clin. Infect. Dis.* **72**, 1879–1887 (2021).
64. Yang, Y.-H. et al. Structural Insights into SraP-Mediated *Staphylococcus aureus* Adhesion to Host Cells. *PLoS Pathog.* **10**, e1004169 (2014).
65. Kukita, K. et al. *Staphylococcus aureus* SasA Is Responsible for Binding to the Salivary Agglutinin gp340, Derived from Human Saliva. *Infect. Immun.* **81**, 1870–1879 (2013).
66. Lees, J. A. et al. Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. *Microb. Genomics* **3**, 1–12 (2017).
67. Schlag, S. et al. Characterization of the Oxygen-Responsive NreABC Regulon of *Staphylococcus aureus*. *J. Bacteriol.* **190**, 7847–7858 (2008).
68. Fuchs, S., Pane, J., Kohler, C., Hecker, M. & Engelmann, S. Anaerobic Gene Expression in *Staphylococcus aureus*. *J. Bacteriol.* **189**, 4275–4289 (2007).
69. Schlag, S., Nerz, C., Birkenstock, T. A., Altenberend, F. & Go, F. Inhibition of *Staphylococcal* Biofilm Formation by Nitrite. *J. Bacteriol.* **189**, 7911–7919 (2007).
70. Lee, E. R., Blount, K. F. & Breaker, R. R. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol.* **6**, 187–194 (2009).
71. Clarke, S. R. et al. Identification of In Vivo-Expressed Antigens of *Staphylococcus aureus* and Their Use in Vaccinations for Protection against Nasal Carriage. *J. Infect. Dis.* **193**, 1098–1108 (2006).
72. Mulcahy, M. E. et al. Nasal Colonisation by *Staphylococcus aureus* Depends upon Clumping Factor B Binding to the Squamous Epithelial Cell Envelope Protein Loricrin. *PLoS Pathog.* **8**, e1003092 (2012).
73. Schaffer, A. C. et al. Immunization with *Staphylococcus aureus* Clumping Factor B, a Major Determinant in Nasal Carriage, Reduces Nasal Colonization in a Murine Model. *Infect. Immun.* **74**, 2145–2153 (2006).
74. Stephens, A. C., Banerjee, S. K. & Richardson, A. R. Specialized phosphate transport is essential for *Staphylococcus aureus* nitric oxide resistance. *mBio* <https://doi.org/10.1128/mbio.02451-23> (2023).
75. Pollitt, E. J. G., West, S. A., Crusz, S. A., Burton-Chellew, M. N. & Diggle, S. P. Cooperation, quorum sensing, and evolution of virulence in *Staphylococcus aureus*. *Infect. Immun.* **82**, 1045–1051 (2014).
76. Stapleton, M. R. et al. Characterization of IsaA and SceD, Two Putative Lytic Transglycosylases of *Staphylococcus aureus*. *J. Bacteriol.* **189**, 7316–7325 (2007).
77. Costa, T. M., Viljoen, A., Towell, A. M., Dufrêne, Y. F. & Geoghegan, J. A. Fibronectin binding protein B binds to loricrin and promotes corneocyte adhesion by *Staphylococcus aureus*. *Nat. Commun.* **13**, 4–5 (2022).
78. Ernst, C. M. et al. Gain-of-Function Mutations in the Phospholipid Flippase MprF Confer Specific Daptomycin Resistance. *mBio* **9**, e01659–18 (2018).
79. Peschel, A. et al. *Staphylococcus aureus* Resistance to Human Defensins and Evasion of Neutrophil Killing via the Novel Virulence Factor MprF Is Based on Modification of Membrane Lipids with L-Lysine. *J. Exp. Med.* **193**, 1067–1076 (2001).
80. Kurokawa, K. et al. Glycoepitopes of *Staphylococcal* Wall Teichoic Acid Govern Complement-mediated Opsonophagocytosis via Human Serum Antibody and Mannose-binding Lectin. *J. Biol. Chem.* **288**, 30956–30968 (2013).
81. Li, X. et al. An accessory wall teichoic acid glycosyltransferase protects *Staphylococcus aureus* from the lytic activity of Podoviridae. *Sci. Rep.* **5**, 17219 (2015).
82. Brown, S. et al. Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. *Proc. Natl. Acad. Sci.* **109**, 18909–18914 (2012).
83. Acker, K. P. et al. Strains of *Staphylococcus aureus* that Colonize and Infect Skin Harbor Mutations in Metabolic Genes. *iScience* **19**, 281–290 (2019).
84. Harkins, C. P. et al. The widespread use of topical antimicrobials enriches for resistance in *Staphylococcus aureus* isolated from patients with atopic dermatitis. *Br. J. Dermatol.* **179**, 951–958 (2018).
85. Riquelme, S. A., Wong Fok Lung, T. & Prince, A. Pulmonary Pathogens Adapt to Immune Signaling Metabolites in the Airway. *Front. Immunol.* **11**, 1–14 (2020).
86. Wildeman, P. et al. Genomic characterization and outcome of prosthetic joint infections caused by *Staphylococcus aureus*. *Sci. Rep.* **10**, 5938 (2020).
87. Lilje, B. et al. Whole-genome sequencing of bloodstream *Staphylococcus aureus* isolates does not distinguish bacteraemia from endocarditis. *Microb. Genomics* **3**, 1–11 (2017).
88. Prijbeliski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinforma.* **70**, e102 (2020).
89. Parkhill, J. et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb. Genomics* **2**, 1–7 (2016).
90. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
92. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
93. Page, A. J., Taylor, B. & Keane, J. A. Multilocus sequence typing by blast from de novo assemblies against PubMLST. *J. Open Source Softw.* **1**, 118 (2016).
94. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
95. Coll, F., Raven, K., Harrison, E. M., Parkhill, J. & Peacock, S. J. *Staphylococcus aureus* core genome coordinates on the ST22 strain HO 5096 0412. https://figshare.com/articles/Staphylococcus_aureus_core_genome_coordinates_on_the_ST22_strain_HO_5096_0412/11627193/2, <https://doi.org/10.6084/m9.figshare.11627193.v2> (2020).
96. Richardson, E. J. et al. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-018-0617-0> (2018).

97. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
98. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
99. Golubchik, T. et al. Within-Host Evolution of *Staphylococcus aureus* during Asymptomatic Carriage. *PLoS ONE* **8**, 1–14 (2013).
100. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evolution* **36**, 2069–2085 (2019).
101. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
102. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. *J. Computational Biol.* **13**, 1028–1040 (2006).
103. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
104. Deatherage, D. E. & Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq. in 165–188 https://doi.org/10.1007/978-1-4939-0554-6_12 (2014).
105. Fuchs, S. et al. Aureo Wiki- The repository of the *Staphylococcus aureus* research and annotation community. *Int. J. Med. Microbiol.* **308**, 558–568 (2018).
106. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2017).
107. The European Committee on Antimicrobial Susceptibility Testing. EUCAST Antimicrobial susceptibility testing. https://www.eucast.org/ast_of_bacteria/.
108. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
109. Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinforma.* **17**, 172 (2016).
110. Wang, B., Zhao, A., Novick, R. P. & Muir, T. W. Activation and Inhibition of the Receptor Histidine Kinase AgrC Occurs through Opposite Helical Transduction Motions. *Mol. Cell* **53**, 929–940 (2014).

Acknowledgements

This publication presents independent research supported by Wellcome grants 201344/Z/16/Z and 204928/Z/16/Z awarded to Francesc Coll. This publication was also supported by the Health Innovation Challenge Fund (WT098600, HICF-T5-342), a parallel funding partnership between the Department of Health and Wellcome Trust. E.M.H. was supported by a UK Research and Innovation (UKRI) Fellowship: MR/S00291X/1. MST was a Wellcome Trust Clinical PhD Fellow at the University of Cambridge. RCM is funded by a Science Foundation Ireland Frontier for the Future Program Award (reference: 21/FFP-A/9704), and a Wellcome Trust Investigator Award (ref: 212258/Z/18/Z). This publication was also supported by Wellcome Grant reference: 220540/Z/20/A, 'Wellcome Sanger Institute Quinquennial Review 2021-2026' and Wellcome

Collaborative Award in Science: 211864/Z/18/Z. This research was supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312*). The views expressed in this publication are those of the author(s) and not necessarily those of the funders.

Author contributions

Conceptualization: F.C., E.M.H.; Data curation: M.T., F.C.; Formal bioinformatic analysis: F.C., M.M.; Funding acquisition: F.C., E.M.H., S.J.P.; Investigation: F.C., E.M.H.; Bioinformatics methodology: F.C., M.M. and D.J.; Laboratory methodology: B.B., K.B. and E.M.H.; Project administration: E.M.H. and S.J.P.; Resources: J.P., E.M.H. and S.J.P.; Supervision: E.M.H., J.A.G., J.P. and S.J.P.; Results contextualization: F.C., E.M.H., R.C.M., T.W., and J.A.G. Validation: B.B., K.B. and R.C.M.; Visualization: FC; Writing – original draft: F.C. and E.M.H.; Writing – review & editing: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-55186-x>.

Correspondence and requests for materials should be addressed to Francesc Coll or Ewan M. Harrison.

Peer review information *Nature Communications* thanks Stephanie Souza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025