Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Limited Concordance Between Contemporaneous Nasopharyngeal Carriage and Invasive Staphylococcus aureus Isolates From The Same Patient among Gambian Children

Sulayman Sowe
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Morrow Minteh
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Horeja A Drammeh
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Matty Faal
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Gorrgi-Sam John
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Paulina Mendy
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Ebrima Faal
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Mustapha Dibbasey
University of Ghana
Grant Makenzie
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Benjamin P Howden
University of Melbourne
Anna Roca
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Ousman Secka
MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine
Lamarana Jallow
University of Ghana
Abdoulie Bojang
Abdoulie.Bojang@lshtm.ac.uk

MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine

Article

Keywords: Staphylococcus aureus, nasopharyngeal carriage, bloodstream infection, invasive disease, whole-genome sequencing, MLST, virulence genes, The Gambia

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11483252/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Staphylococcus aureus nasopharyngeal colonisation is a major risk factor for invasive disease and has traditionally been regarded as the principal endogenous reservoir from which bloodstream infections arise. However, the genomic relationship between contemporaneously sampled carriage and invasive isolates remains incompletely understood, particularly in sub-Saharan Africa, where paired genomic studies are scarce. In this study, we performed whole-genome sequencing of paired bloodstream (BC) and contemporaneously collected nasopharyngeal (NPS) *S. aureus* isolates from 35 Gambian children with invasive disease. To improve representation of the carriage population, up to eight morphologically distinct colonies were analysed from each nasopharyngeal specimen. Paired isolates were compared for multilocus sequence type (MLST) concordance, clonal complex distribution, antimicrobial resistance determinants, and virulence-gene composition. Our results showed that only 9 of 35 (25.7%) patients had at least one nasopharyngeal isolate sharing the same sequence type as the paired bloodstream isolate, indicating limited patient-level concordance. Despite this, bloodstream and nasopharyngeal isolates displayed remarkably similar genetic structures with ST152/CC152 and ST15/CC15 dominating both compartments. Within the ST-matched cohort, most virulence determinants were shared between the matched isolates; however, certain genes, including *cap operon*, *cna*, and *hld*, were more frequently identified among bloodstream isolates, whereas *chp* gene was more prevalent among nasopharyngeal isolates. Antimicrobial resistance (AMR) gene profiles were broadly similar, with no clear evidence of differential AMR-gene prevalence. Contemporaneously sampled nasopharyngeal carriage frequently failed to capture the *S. aureus* lineage recovered from blood, despite both compartments sharing the same dominant circulating lineages. These findings suggest that patient-level carriage–bloodstream concordance cannot be inferred from population-level lineage similarity and highlight the limitations of single-site carriage sampling for investigating the origins of invasive disease. Higher-resolution genomic analyses and longitudinal multi-site sampling will be important to determine the evolutionary and ecological relationships between colonising and invasive *S. aureus* populations.

Introduction

Staphylococcus aureus is both a common human commensal and one of the world's most consequential bacterial pathogens, causing diseases ranging from superficial skin and soft-tissue infections to life-threatening pneumonia, infective endocarditis, bacteraemia, osteomyelitis, and sepsis. Globally, it is the leading bacterial cause of bloodstream infection and was associated with an estimated 1.1 million deaths in 2019, making it the only bacterial pathogen linked to more than one million deaths worldwide [1]. Despite decades of research, no licensed vaccine is available, and multiple vaccine candidates have failed to demonstrate effective protection against invasive disease [2,3]. Together with the continued emergence of antimicrobial resistance, these challenges have shifted attention towards understanding the earliest stages of *S. aureus* pathogenesis, particularly the transition from asymptomatic colonisation to invasive infection. Colonisation, particularly within the upper respiratory tract, is a well-established risk factor for subsequent infection because carried populations may provide an endogenous reservoir from

which invasive strains emerge. This relationship has shaped current prevention strategies, including nasal screening and decolonisation, which are largely based on the premise that the strain causing invasive disease originates from the patient's colonising population.

Advances in whole-genome sequencing have refined our understanding of this carriage-to-disease relationship while also revealing its complexity. Closely related carriage and invasive isolates have been recovered from the same patients, sometimes with little detectable genomic or phenotypic adaptation [4,5]. However, other studies have documented within-host diversification, strain replacement, and adaptive mutations affecting regulatory and host-interaction pathways during progression to disease [6,7].

At the population level, carriage and invasive isolates frequently occupy overlapping clonal backgrounds and share much of their classical virulence repertoire [8–10]. Together, these findings indicate that invasive disease is not restricted to a distinct pathogenic lineage and that lineage concordance alone cannot fully explain how colonising populations progress to invasive infection. Furthermore, whether carriage and invasive isolates sharing the same MLST background differ in virulence-associated gene content within individual patients remains insufficiently understood. In comparison to a population-level analysis, a within-patient analysis reduces some confounding by lineage and host. We sought to understand why some carriage strains become invasive while others don't, and what antimicrobial resistance and virulence genes differences occur during the transition from colonisation to invasion.

An additional challenge is that most genomic comparisons have relied on a single colony recovered from a single carriage site, an approach that may underestimate the diversity of *S. aureus* populations within individual hosts. Increasing evidence suggests that carriage is frequently multiclonal and dynamic, with co-existing lineages, temporal turnover, and ecological interactions that may not be adequately captured by conventional sampling strategies [11–13]. As a result, whether a contemporaneously sampled carriage isolate accurately represents the bacterial lineage responsible for invasive disease remains unresolved.

This uncertainty is particularly important in sub-Saharan Africa, where *S. aureus* carriage is common during early life but paired genomic studies remain scarce. In The Gambia, nasopharyngeal carriage has been reported in approximately 50–63% of newborns and infants [14]. More recently, a paired genomic study from The Gambia identified matching clinical and oropharyngeal sequence types in four of five neonates, providing important evidence that upper-respiratory carriage can serve as a reservoir for invasive infection [15]. However, the small number of paired cases and limited representation of within-host carriage diversity leave unresolved whether contemporaneously sampled nasopharyngeal populations consistently capture the invasive lineage across a larger paediatric population. Addressing this question is essential for interpreting carriage surveillance, understanding the reservoirs of invasive disease, and informing future prevention strategies.

In this study, we analysed paired bloodstream and contemporaneously collected nasopharyngeal *S. aureus* isolates from 35 Gambian children with invasive disease. Unlike conventional single-colony comparisons, we examined multiple nasopharyngeal colonies from each patient to better capture within-

host carriage diversity. By comparing the lineage concordance, population structure, virulence-gene composition and antimicrobial resistance determinants of paired isolates, we sought to determine whether contemporaneously sampled nasopharyngeal carriage accurately represents the *S. aureus* population associated with invasive disease.

Results

Study population

We sequenced the genomes of 142 *S. aureus* isolates. These samples were collected from 35 patients suffering from invasive *S. aureus* infection. One BC isolate from each patient totalling 35 isolates. From the same 35 patients, an additional NPS was also collected and up to 8 distinct colonies of *S. aureus* were picked giving a total of 107 NPS isolates. This cohort was heavily skewed towards young infants (median age:9.0 months). Cases peaked during the high transmission rainy season (62.9% vs 37.1% dry season) (**Table 1**).

Primary finding

A direct individual level MLST comparison revealed low concordance (25.7%; 9/35). Only nine patients had invasive disease caused by a strain that was found in the sampled nasopharynx (**Figure 1**).

Genetic lineages

To explain this observed low concordance, we first looked at the lineages of the isolates to examine whether the unmatched BC isolates clustered in specific STs/CCs which would infer that the invasive strains were derived from a different genetic pool to the NPS isolates. Despite individual level mismatches, the unmatched BC isolates do not cluster into a unique, invasive-only group of STs. Instead, they are dominated by the exact same lineages found in the community carriage pool, specifically ST152/CC152 (22.9%) and ST15/CC15 (17.1%). Furthermore, Fisher's exact testing confirms no significant phylogenetic divergence between the two groups, suggesting that the unmatched blood isolates are not a distinct taxonomic group (**Table 2**) and are likely drawn from the nasopharyngeal isolates.

Virulence profiles

As increased virulence is often associated with specific genetic profiles, we examined the genetic content for virulence factors between BC and NPS isolates via generalised linear mixed-effects modeling. The vast majority of the 30 evaluated virulence factors showed profound population-level symmetry, with core pathogenic determinants such as alpha-hemolysin (*hla/hly*), core exoenzymes (*aur, nuc, lip*), and the iron-regulated surface determinants (*isd* complex) exhibiting near-universal conservation and no statistical divergence between compartments ($p = 1.00$ or Invariant). However, we observed significant virulence gene associations with invasive disease particularly, the capsule operon (*cap operon*; 42.9% BC vs 15.0% NPS, OR 11.44, $p < 0.001$), the collagen-binding adhesin (*cna*; 80.0% BC

vs 53.3% NPS, OR 4.46, p = 0.006), and staphylokinase (*sak*; 82.9% BC vs 65.4% NPS, OR 3.97, p = 0.031) and significant virulence gene associations with colonisation particularly, the chemotaxis inhibitory protein (*chp* 60.0% BC vs 79.4% NPS, OR 0.21, p = 0.005) at the population level (**Supplementary Table S1**).

Separately, a virulence genes analysis table between the BC and NPS isolate(s) within the nine patients that had the same MLST background revealed a similar result (**Supplementary Table S2; Fig. 2**). The *cap operon*, *cna* and *hld* were significantly associated with invasiveness whiles *chp* was significantly associated with colonisation. The mean virulence genes prevalence within the ST-matched population between the BC and NPS is not significantly different (p = 0.20716) (**Supplementary Fig. S2**). However, the mean prevalence of target virulence gene associated with invasive and carriage were significantly different. In the nasopharynx, the prevalence of (*chp+*) gene associated with carriage was 87.0%, whilst the prevalence of invasive associated genes (*cap+/can+/hld+*) were only 12.5%. In the blood, the carriage associated gene prevalence dropped to 56.0%, whilst the invasive associated genes surges to 33.0%.

Antimicrobial resistance profiles

AMR gene profiles were broadly similar between BC and NPS isolates within all 35 patients and within the ST-matched patients (**Supplementary Tables S3 and S4; Supplementary Fig. S1**). No statistically significant associations were identified for the commonly detected genes, including *blaZ* and *dfrG*. Although *tet(K)* was more frequent among BC isolates, this association was not statistically significant. Estimates for rare or absent genes were not interpretable because of separation or invariant prevalence.

Phylogenetic relationships between BC and NPS isolates

In this phylogenetic analysis, several patients had relatively short minimum patristic distances between their BC and NPS isolates, particularly among ST-concordant pairs, whereas ST-mismatched pairs generally showed larger BC–NPS distances. These relationships were assessed using patristic distances between isolate tips rather than MLST alone (**Fig. 3; Supplementary Table S5**).

ST-concordant patients

Seven of the 18 patients in the phylogenetic subset were classified as ST-concordant according to the MLST data. In these patients, the closest BC–NPS patristic distances in ST-matched patients were generally lower than those observed in ST-mismatched patients. The minimum patristic distances among the seven labelled ST-concordant patients ranged from 0.002522679 to 0.042976313.

The smallest minimum distances were observed for patients 30333, 39725, 33113, 36456, 30155 and 33440. Patient 22211 had a larger minimum distance of 0.042976313 despite being classified as ST-concordant.

ST-mismatched patients

Among the 11 patients classified as ST-mismatched, the minimum BC–NPS patristic distances ranged from (0.081529421) to (0.271135342). These values were generally higher than those observed among the ST-concordant patients. The closest NPS isolate in the ST-mismatched group was observed for patient 6530, while the largest minimum distances were observed for 4034, 33153, and 475369.

Discussion

This study provides a patient-level genomic assessment of the relationship between invasive and contemporaneously sampled nasopharyngeal *Staphylococcus aureus* collected from the same patient among Gambian children. The principal finding was the low MLST concordance between paired compartments: only 9 of 35 patients had at least one nasopharyngeal isolate sharing the same sequence type to the corresponding bloodstream isolate. Therefore, in approximately three-quarters of patients, the invasive lineage was not recovered from the nasopharyngeal population sampled at the time of invasive disease. This finding indicates that a contemporaneous nasopharyngeal sample often provides an incomplete representation of the *S. aureus* lineage detected in blood.

Previous paired studies have frequently reported much higher carriage–infection concordance, including [5], [4], and the earlier Gambian neonatal study [15]. However, many of these studies analysed one carriage colony per sample or included relatively small numbers of paired patients. By examining up to eight colonies from each NPS specimen, the present study sampled a broader section of the carriage population. This is consistent with [13], who demonstrated extensive multi-strain carriage and intra-host diversity across upper-respiratory sites, and with [12], who showed that incomplete sampling of within-host diversity can alter genomic relatedness inference.

A striking feature of the results was the contrast between low patient-level concordance and substantial population-level overlap. ST152/CC152 and ST15/CC15 were dominant in both BC and NPS populations, and no common sequence type clearly defined an invasive-only population.

This population-level overlap agrees with studies showing that carriage and disease isolates are not separated into distinct evolutionary populations [8–10]. [8] found that clinical and nasal isolates shared evolutionary lineages, while [9] showed substantial overlap in clonal complexes among commensal and invasive isolates. More recent population-genomic analyses have reached a similar conclusion. [16] found that carriage and disease isolates from the same communities were drawn from overlapping genomic populations rather than from clearly separated pathogenic and commensal lineages. Together with the present findings, this suggests that invasive disease is not restricted to a discrete group of inherently invasive STs. Instead, circulating lineages appear capable of occupying both carriage and disease states, with patient-level outcome influenced by factors beyond MLST background alone.

The virulence-gene analysis further supports this interpretation. Most classical virulence determinants were widely conserved between BC and NPS isolates, including genes involved in iron acquisition, exoenzyme production, adhesion, and global regulation. This broad similarity is consistent with [4], who found no acquisition of new virulence genes during progression from nasal carriage to bacteraemia, and with [5], who reported limited genomic and phenotypic differentiation between colonising and invasive isolate pairs. It also agrees with [17], who found that carriage and bloodstream populations could not be consistently distinguished by virulence traits, emphasizing that carriage isolates may already possess substantial pathogenic potential.

However, the virulence analysis also identified targeted differences within this shared genomic background. The *capsule operon*, *cna*, and *hld* were more prevalent among bloodstream isolates, whereas *chp* was more frequent among NPS isolates. These findings suggest that the transition from carriage-associated populations to invasive disease may involve differences in selected capsule, adhesion, and toxin functions rather than wholesale acquisition of a distinct virulence repertoire. The differential prevalence of *cna* is biologically plausible because collagen binding may facilitate tissue adherence and dissemination, while capsule production may enhance resistance to phagocytic clearance. Similarly, *hld* may damage host cell membranes. [9] also identified selected virulence determinants associated with invasive isolates, and [18] found infection-associated genomic signatures involving adhesins, toxins, capsule-related genes, and immune-modulatory functions. The present results suggest that differences in accessory virulence-associated gene content may contribute to variation in invasive potential among isolates within shared lineages.

These associations should nevertheless be regarded as preliminary rather than definitive markers of invasiveness. Gene presence does not establish whether the locus is intact, expressed, or functionally active. The finding that most virulence genes were shared indicates that invasion is unlikely to be explained by gene content alone. [19] and [7] showed that short-term adaptive changes affecting regulatory and host-interaction pathways can accompany the transition from carriage to disease without major changes in accessory-gene content. Thus, sequence variation, gene integrity, and regulation may be more informative than presence–absence analysis alone.

Whole-genome phylogenetic analysis provided additional resolution beyond MLST in the 18-patient subset. Several ST-concordant BC–NPS pairs clustered closely, with comparatively small patristic distances, supporting the presence of closely related carriage and invasive populations in some patients. In contrast, ST-mismatched patients generally had larger minimum BC–NPS patristic distances, consistent with carriage and bloodstream isolates representing more divergent members of the broader *S. aureus* population.

These findings should nevertheless be interpreted cautiously. The phylogenetic analysis included only 18 of the 35 patients, and patristic distance is dependent on the selected alignment and phylogenetic model. In addition, the analysis does not establish the direction or timing of transmission, nor does it determine whether the BC strain originated from the sampled nasopharynx. Longitudinal sampling and

broader anatomical-site sampling would be required to distinguish persistent carriage, strain replacement, transient colonisation, and acquisition from an unsampled reservoir.

The combined findings suggest several possible explanations for the low within-patient concordance. The invasive lineage may have been present earlier in the nasopharynx but replaced before sampling, consistent with longitudinal evidence that carriage populations evolve and change over time [4,19]. It may also have been present at low abundance or located in another unsampled reservoir. [11] showed that *S. aureus* colonisation exists along an abundance-dependent and microbiome-structured continuum rather than as a simple binary state, providing a biological basis for transient or low-level carriage that may evade conventional culture. Long-term carriage studies have also demonstrated mutation accumulation, population turnover, and strain replacement within individual hosts [6]. Therefore, the absence of the bloodstream ST from the contemporaneous NPS sample may reflect the temporal and ecological instability of carriage rather than complete absence of an endogenous reservoir.

Host factors are also likely to contribute to progression from carriage or acquisition to invasive disease. Although BC and NPS isolates in the ST-matched cohort shared MLST backgrounds, BC isolates showed higher prevalence of the *cap operon*, *cna*, and *hld*, whereas *chp* was more prevalent among NPS isolates. Thus, invasive disease may reflect an interaction between modest differences in virulence-associated gene content and host conditions, rather than the emergence of a wholly distinct invasive lineage. Barrier disruption, immune immaturity, underlying illness, or medical interventions have been identified as an important event in this disease continuum [5,17]. In the present paediatric cohort, the young median age may be especially relevant because immature host defenses could increase susceptibility to bloodstream invasion by lineages that are also common in asymptomatic carriage.

Overall, our findings reveal a marked disconnect between population-level and patient-level relationships in *S. aureus*. Bloodstream and nasopharyngeal populations shared the same dominant circulating lineages, yet the invasive ST was recovered from contemporaneous nasopharyngeal carriage from the same patient in only a minority of children. This distinction is important for interpreting carriage surveillance and suggests that a single contemporaneous nasopharyngeal sample may not reliably represent the *S. aureus* lineage associated with invasive disease.

Materials and methods

Sample sources and selection

This retrospective study drew on participants from two main sources: the Pneumococcal Vaccine Schedules (PVS) study and the Pneumococcal Surveillance Project (PSP).

The PVS study was a large cluster-randomized, non-inferiority trial comparing a reduced two-dose infant pneumococcal conjugate vaccine schedule (1 + 1) with the standard three-dose schedule (3 + 0). During

the trial, participants suspected of invasive bacterial disease had blood collected for culture to identify the causative organism, and a subset also had a nasopharyngeal swab collected [20].

The PSP was a population-based surveillance program for invasive pneumococcal disease among individuals aged ≥ 2 months residing within the Basse Health and Demographic Surveillance System (BHDSS) in the Upper River Region, The Gambia. This was conducted between May 2008 and December 2014. In addition to pneumococci, the PSP also isolated other bacterial pathogens from blood cultures, and similarly nasopharyngeal swab was collected from a subset of participants [21].

Only participants sampled for both their *S. aureus* positive blood culture and paired nasopharyngeal swab in both studies have been included in this study (**Isolates Flowchart**). The samples from the two groups are referred to as BC (blood culture) and NPS (nasopharyngeal swab) isolates, respectively. Demographic variables, including patient age, sex, and date of isolation (stratified into distinct dry and rainy seasons), were tabulated to characterise the clinical cohort (Table 1).

A total of 142 *S. aureus* isolates were included in this study.

Isolates Flowchart

Microbial culture

Frozen vials containing BC or NPS in skim milk, tryptone, glucose, glycerol (STGG), were retrieved from −80°C freezers in MRCG Biobank and transported to the Research Microbiology laboratory for processing. The vials were allowed to thaw briefly at room temperature before further processing.

A 10µl loop full of retrieved BC samples were initially streaked on Mannitol Salt Agar (MSA) plates to obtain discrete colonies and incubated overnight at 37°C. For each MSA plate, a single, yellow-pigmented isolated colony was sub-cultured onto Blood Agar (BA) plate and incubated overnight at 37°C to obtain pure colonies.

NPS samples were vortexed briefly before 100 µl was dispensed onto MSA plates, streaked and incubated at 37°C overnight. For each NPS sample, up to a maximum of eight distinct, yellow-pigmented colonies were sub-cultured on BA plates overnight to fully capture within-host nasal carriage dynamics.

The BA plates of both specimen types were examined for β-hemolysis and confirmed with the BactiStaph latex agglutination test (Remel™) according to the manufacturer's instructions. Only positively confirmed isolates were selected for downstream processing.

Cell Lysis and DNA extraction

A quarter of a 10 µl loop of each BactiStaph-positive isolate was transferred into 500 µl of normal saline in 2 mL Sarstedt tubes to prepare bacterial suspensions for cell lysis and DNA extraction. The

suspensions were briefly vortexed and centrifuged at 7050 rpm for 10 minutes at room temperature. The supernatant was carefully discarded, ensuring the pellets remained undisturbed.

To each Sarstedt tube, 160 µl of Buffer P1 and 20 µl of lysozyme (100 mg/mL) were added. The mixture was incubated at 37°C for 2 hours in a thermoshaker. Subsequently, 20 µl of Proteinase K was added, mixed gently by tapping, and incubated overnight at 56°C. After overnight incubation, DNA extraction was performed using the QIAsymphony SP instrument with the DSP DNA Mini Kit, following the “Tissue HC 200 DSP” protocol and an elution volume of 100 µl. DNA quantification was done using both a Nanodrop spectrophotometer and a Qubit fluorometer. The Nanodrop was initially used to determine DNA concentration, purity, and contamination, while the Qubit provided specific DNA quantification for downstream normalization and library preparation.

Library preparation and genomic sequencing

Library preparation was performed as per the Native Barcoding Kit 96 V14 (SQK-NBD114.96) protocol from Oxford Nanopore Technologies (ONT). Up to 96 samples were normalized per library. In a clean 96-well plate, 400 ng of DNA was aliquoted per sample and adjusted to a final volume of 11 µl with nuclease-free water. Samples were gently mixed by pipetting and briefly centrifuged to collect the contents at the bottom of the wells.

We then carried out DNA repair and end-prep. In these steps, the damaged DNA is repaired, and the fragment ends standardized to ensure efficient and accurate adapter ligation. This was followed by native barcode ligation involving the attachment of unique barcode adapters to the end-prepped DNA without PCR amplification, allowing multiple samples to be multiplexed in a single ONT sequencing run. Afterwards, the barcoded DNA from all the samples were pooled into a single DNA LoBind tube, and 1µl qubit fluorometer quantification was done prior to Adapter ligation.

During Adapter ligation, sequencing adapters containing motor proteins were attached to the ends of the barcoded DNA fragments, enabling controlled translocation through the nanopores and efficient sequencing. The adapter-ligated DNA was then cleaned using AMPure XP beads and Short Fragment Buffer to remove excess adapters, enzymes, salts, and buffers. Finally, 25 µl of the purified DNA library was eluted into a clean 1.5 mL DNA LoBind tube, and 1 µl was quantified using a Qubit fluorometer.

The average fragment size of the final library was determined using an Agilent TapeStation with the D5000 sample buffer and ladder as per the manufacturer’s instructions. Based on the measured fragment size and Qubit concentration, 50 fmol of library was calculated and loaded onto the PromethION 24 ONT platform to generate high-yield simplex data.

Reads were basecalled using MinKNOW (version 6.8.7, Oxford Nanopore Technologies) with the super accuracy model. MinKNOW collects sequencing data, basecalls, and demultiplexes in real time. Each sequencing run was done for up to 72hrs with the R10.4 Flow cells. EPI2ME workflows (wf-bacterial-genomes workflow v1.4.5 ONT) were used for quality control, species identification (MLST), and AMR

gene identification (ResFinder), and virulence annotations (Prokka). EPI2ME workflows were employed with a Q-score ≥ 10.

Molecular Characterization and Virulence Profiling

All samples underwent Multi-Locus Sequence Typing (MLST) to determine their Sequence Type (ST) and assign them to corresponding Clonal Complexes (CC). Isolates were screened for a high-density panel of 30 vital *S. aureus* accessory genes distributed across seven functional categories: Adhesion (e.g., *clfA, cna*), Biofilm production (ica, bap), Iron acquisition (isd), Immune evasion (*sak, spa*), Toxins (*hla, tst, psm*), Quorum sensing (*agr, sar*), and Exoenzymes (*aur, hysA*) demonstrating profound functional symmetry. Additionally, genotypic profiles for antimicrobial resistance (AMR) were documented.

Data analysis

Odds Ratios (OR) and 95% Confidence Intervals (CI) were calculated using the broom and stats packages in R. All analyses were performed using R statistical software (v4.5.3; R Core Team 2026) and Microsoft excel (Microsoft Corporation, 2024).

To avoid treating multiple NPS colonies from the same participant as independent observations, analyses were conducted at two levels. Isolate-level frequencies were reported descriptively to summarize laboratory recovery. For comparative analyses, NPS colonies were collapsed to a patient-level binary variable for each gene: a gene was classified as present when detected in at least one colony from that participant and absent when undetected in all colonies. BC and aggregated NPS profiles were then compared within participants using paired analyses.

Phylogenetic analysis

Whole-genome sequence data was available for isolates from 18 patients. Following quality-controlled variant calling, high-quality PASS variants were identified and combined across isolates to generate a multiple sequence alignment of variable genomic sites. A maximum-likelihood phylogeny was reconstructed using IQ-TREE under a GTR-based model incorporating invariant sites and gamma-distributed rate variation. Branch support was assessed using SH-aLRT and ultrafast bootstrap values.

Six non-*S. aureus Staphylococcus* isolates forming a distinct monophyletic clade were used as the biological outgroup for rooting. Because inclusion of this outgroup produced an exceptionally long branch that compressed the ingroup, the outgroup isolates were removed from the principal visualization after tree rooting; branch lengths and the root placement were retained for interpretation.

Tree metadata were used to annotate specimen type, patient identifier, ST, and CC. Within-patient relatedness was summarized using patristic distance between BC and NPS isolate tips. For each patient, we reported the minimum, median, and maximum BC–NPS patristic distance. The NPS isolate with the smallest patristic distance to the BC isolate was identified as the closest NPS isolate.

Declarations

Ethics statement. The study was reviewed and approved by the Gambia Government/MRC Joint Ethics Committee (SCC No. 31142) and the London School of Hygiene & Tropical Medicine Research Ethics Committee (LSHTM Ethics Ref. 31142). Samples analysed in the present study were collected as part of the Pneumococcal Surveillance Project (PSP) and Pneumococcal Vaccine Schedules (PVS) study. Informed consent for participation and sample collection was obtained from the parents or legal guardians of participating children in accordance with the respective study protocols. All procedures were conducted in accordance with the relevant institutional guidelines and regulations.

Data availability.

Data Availability

The BioProject associated with this study has been registered with the National Center for Biotechnology Information (NCBI) under accession number **PRJNA1544551**. The raw Oxford Nanopore sequencing reads are currently undergoing submission to the NCBI Sequence Read Archive (SRA) under submission ID **SUB16591116**. The SRA accession numbers and reviewer-access information will be provided once the submission has been processed by NCBI. The R scripts and other analysis code used in this study are available at GitHub: https://github.com/sulayman-sowe/SAUREUS-MANUSCRIPT_03-FIGURES.

ACKNOWLEDGMENTS

This work was supported by the Gates Foundation (Investment ID INV-007175; formerly OPP1191735) through the African Postdoctoral Training Initiative (APTI). APTI is implemented by the African Academy of Sciences in partnership with the U.S. National Institutes of Health (NIH), where grantees receive mentorship during the first two years of the fellowship. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the views of the Gates Foundation, the African Academy of Sciences, or the U.S. National Institutes of Health.

AUTHOR CONTRIBUTIONS

A.B.: Conceptualization, Investigation, Formal analysis, Writing – original draft. B.P.H.: Conceptualization. A.R.: Conceptualization. S.S.: Investigation, Formal analysis, Writing – original draft. M.M.: Investigation. H.A.D.: Investigation. M.F.: Investigation. P.M.: Investigation. G.S.J. Formal analysis.: E.F.: Investigation. M.D.: Formal analysis. G.M.: Conceptualization. O.S.: Conceptualization. L.J.: Formal analysis, Writing – original draft. All authors contributed to the review and revision of the manuscript and approved the final version.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

References

1. Ikuta, K. S. *et al.* Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **400**, 2221-2248, doi:10.1016/S0140-6736(22)02185-7 (2022).
2. Fowler, V. G., Jr. & Proctor, R. A. Where does a Staphylococcus aureus vaccine stand? *Clinical Microbiology and Infection* **20**, 66-75, doi:10.1111/1469-0691.12570 (2014).
3. Howden, B. P. *et al.* Staphylococcus aureus host interactions and adaptation. *Nature Reviews Microbiology* **21**, 380-395, doi:10.1038/s41579-023-00852-y (2023).
4. 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**, e0195860, doi:10.1371/journal.pone.0195860 (2018).
5. Räz, A. K. *et al.* Limited Adaptation of Staphylococcus aureus during Transition from Colonization to Invasive Infection. *Microbiol Spectr* **11**, e0259021, doi:10.1128/spectrum.02590-21 (2023).
6. Larsen, T. G., Samaniego Castruita, J. A., Worning, P., Westh, H. & Bartels, M. D. Within-host genomic evolution of methicillin-resistant Staphylococcus aureus in long-term carriers. *Applied Microbiology and Biotechnology* **108**, 95, doi:10.1007/s00253-023-12932-3 (2024).
7. Young, B. C. *et al.* Severe infections emerge from commensal bacteria by adaptive evolution. *eLife* **6**, e30637, doi:10.7554/eLife.30637 (2017).
8. Lamers, R. P., Stinnett, J. W., Muthukrishnan, G., Parkinson, C. L. & Cole, A. M. Evolutionary analyses of Staphylococcus aureus identify genetic relationships between nasal carriage and clinical isolates. *PLoS One* **6**, e16426, doi:10.1371/journal.pone.0016426 (2011).
9. Rasmussen, G., Monecke, S., Ehricht, R. & Söderquist, B. Prevalence of clonal complexes and virulence genes among commensal and invasive Staphylococcus aureus isolates in Sweden. *PLoS One* **8**, e77477, doi:10.1371/journal.pone.0077477 (2013).
10. Roe, C. *et al.* Genomic analyses of Staphylococcus aureus clonal complex 45 isolates does not distinguish nasal carriage from bacteraemia. *Microb Genom* **6**, doi:10.1099/mgen.0.000403 (2020).
11. Aggarwal, D. *et al.* Large-scale characterisation of the nasal microbiome redefines Staphylococcus aureus colonisation status. *Nature Communications* **16**, 10415, doi:10.1038/s41467-025-66564-4 (2025).
12. Hall, M. D. *et al.* Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *eLife* **8**, e46402, doi:10.7554/eLife.46402 (2019).
13. Webb, J. *et al.* Multi-strain carriage and intrahost diversity of Staphylococcus aureus among Indigenous adults in the USA. *Microbial Genomics* **11**, doi:https://doi.org/10.1099/mgen.0.001367 (2025).
14. Ebruke, C. *et al.* High genetic diversity of Staphylococcus aureus strains colonising the nasopharynx of Gambian villagers before widespread use of pneumococcal conjugate vaccines. *BMC Microbiol* **16**, 38, doi:10.1186/s12866-016-0661-3 (2016).
15. Bojang, A. *et al.* Genomic approach to determine sources of neonatal Staphylococcus aureus infection from carriage in the Gambia. *BMC Infect Dis* **24**, 941, doi:10.1186/s12879-024-09837-5

(2024).

16. Cella, E. *et al.* Population genomic variation in Staphylococcus aureus isolates from carriage and disease in Indigenous communities in the Southwest United States. *Microbial Genomics* **12**, doi:https://doi.org/10.1099/mgen.0.001666 (2026).
17. Raineri, E. J. M. *et al.* Staphylococcus aureus populations from the gut and the blood are not distinguished by virulence traits—a critical role of host barrier integrity. *Microbiome* **10**, 239, doi:10.1186/s40168-022-01419-4 (2022).
18. Chen, J. *et al.* Genome-based model for differentiating between infection and carriage Staphylococcus aureus. *Microbiol Spectr* **12**, e0049324, doi:10.1128/spectrum.00493-24 (2024).
19. Young, B. C. *et al.* Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. *Proc Natl Acad Sci U S A* **109**, 4550-4555, doi:10.1073/pnas.1113219109 (2012).
20. Mackenzie, G. A. *et al.* Pneumococcal vaccine schedules (PVS) study: a cluster-randomised, non-inferiority trial of an alternative versus standard schedule for pneumococcal conjugate vaccination-statistical analysis plan. *Trials* **23**, 1058, doi:10.1186/s13063-022-06900-x (2022).
21. Mackenzie, G. A. *et al.* Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. *Lancet Infect Dis* **16**, 703-711, doi:10.1016/s1473-3099(16)00054-2 (2016).

Tables

Tables 1 and 2 are available in the supplementary files section

Figures

Figure 1:

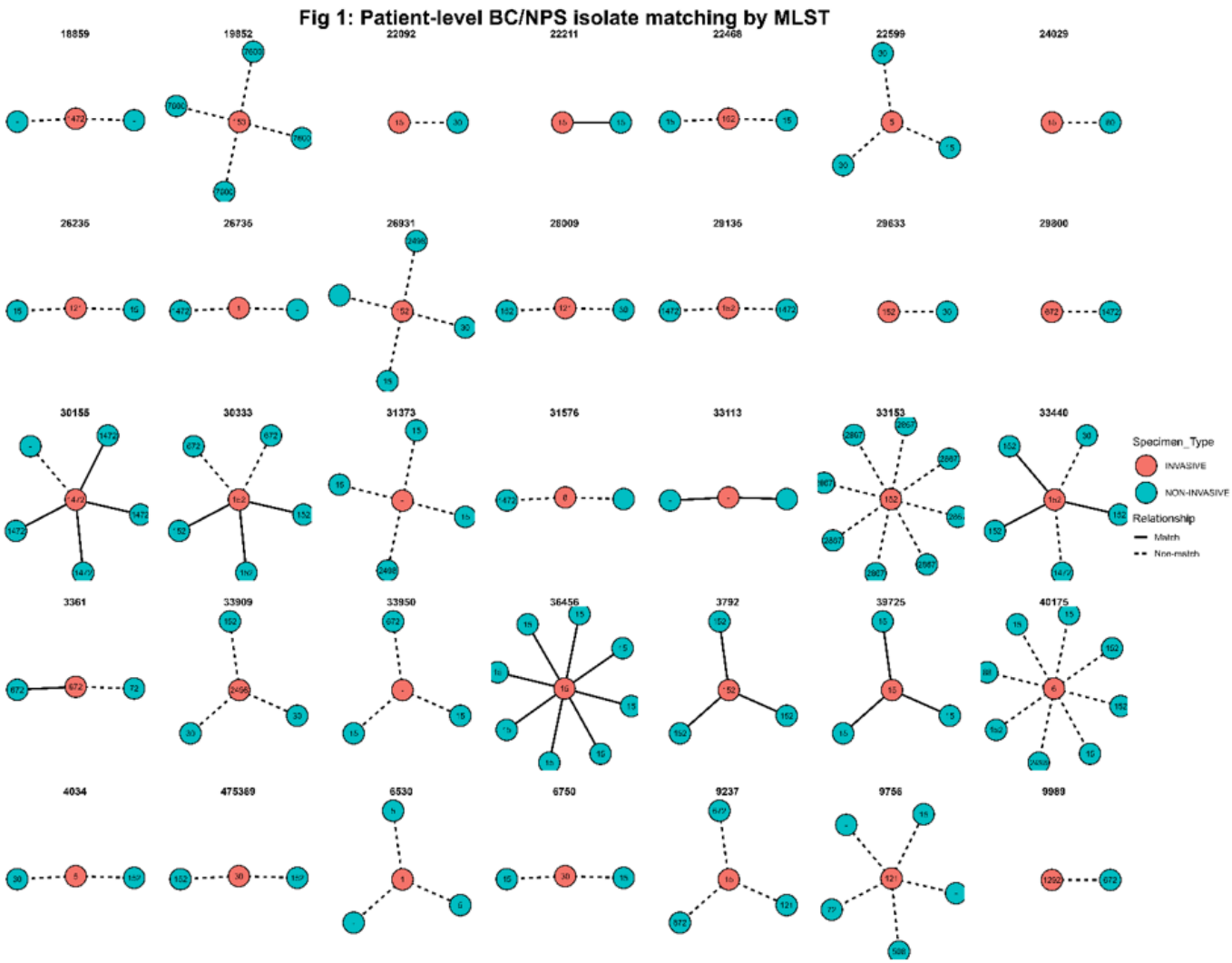


Figure 1

Patient-Level BC/NPS MLST Concordance

This network diagram illustrates patient-level blood culture (BC) and nasopharyngeal swab (NPS) isolate matching using Multi-Locus Sequence Typing (MLST). Each sub-panel represents an individual participant tracked by their unique Study ID. Nodes are color-coded by sample origin, tracking invasive isolates (red nodes) and non-invasive carriage isolates (teal nodes), with sequence types (STs) labeled inside. Connecting vector patterns denote the specific typing relationship, using solid lines for exact ST matches and dashed lines for ST mismatches, mapping out the localized population architecture and strain sharing within individual hosts.

Figure 2:

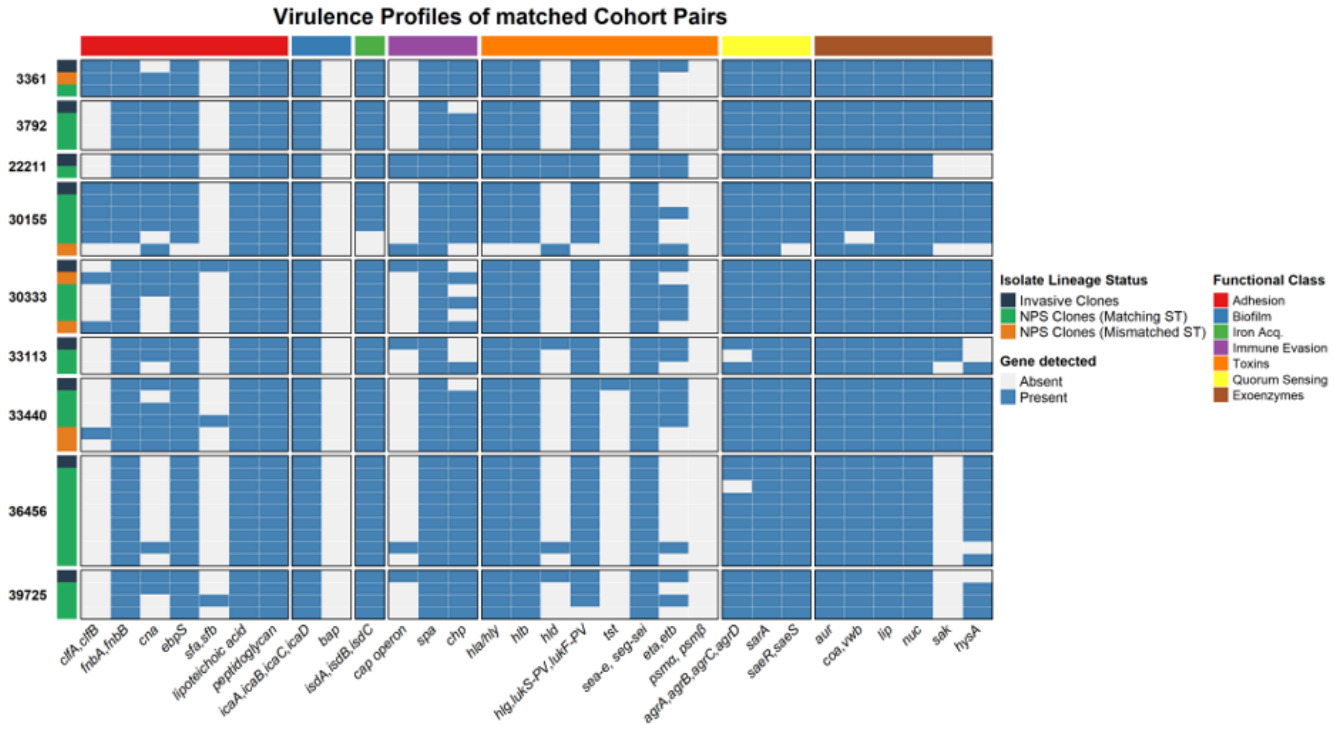


Figure 2

Virulence Profiles in ST-Matched Patients

This heatmap reveals the virulence gene profiles of ST-matched patients. The horizontal columns represent specific virulence factors grouped under their physiological functional classes, including Adhesion, Biofilm formation, Iron acquisition, Immune evasion, Toxins, Quorum sensing, and Exoenzymes. Individual rows organize the clinical isolates by patient ID, with color blocks mapping out the lineage concordance status (invasive vs. matching or mismatched carriage clones). Filled blue tiles denote the detection of a target virulence gene, highlighting specific conservation or missing components between systemic and mucosal niches.

Figure 3

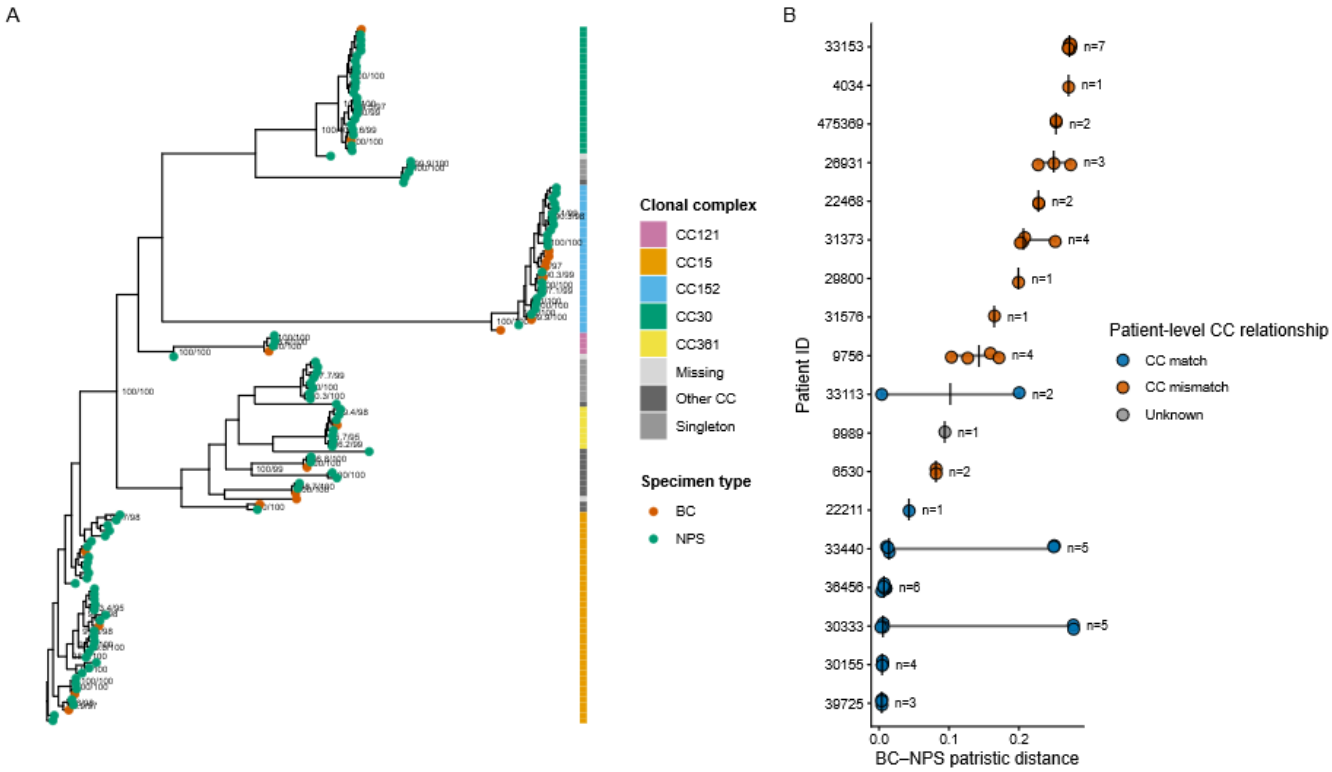


Figure 3

Core-Genome Phylogeny and Within-Patient Patristic Distances

This multi-panel analysis establishes the evolutionary relationship of the *S. aureus*cohort. Panel (A) displays a maximum-likelihood core-genome phylogenetic tree, with branch leaf nodes colored by specimen type (BC vs. NPS) and flanked by colored side-bars defining major Clonal Complexes (CC). Panel (B) details the within-patient phylogenetic relatedness by plotting the absolute core-genome BC–NPS patristic distance for each Patient ID. The horizontal vector length spans individual multi-isolate samples (n), while point markers denote whether the systemic disease clone represents a CC match (blue circles) or a distinct colonizing CC mismatch (orange circles) within the host.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryinformationFINAL.docx
- Tables.docx