

Contrasting outcomes of cross-species transmission in *Staphylococcus aureus* ST30

Agnes Marie Figueiredo

agnes@micro.ufrj.br

Universidade Federal do Rio de Janeiro

Matheus Assis Côrtes Esteves

<https://orcid.org/0000-0003-2282-8581>

Mariana Carvalho

Universidade Federal do Rio de Janeiro

Caroline Martini

Universidade Federal do Rio de Janeiro <https://orcid.org/0000-0001-8637-6110>

Luís Guilherme Longo

Universidade Federal do Rio de Janeiro

Laura Oliveira

Federal University of Rio de Janeiro <https://orcid.org/0000-0002-7162-2013>

Tatiana Pinto

Federal University of Rio de Janeiro <https://orcid.org/0000-0002-7878-9377>

Article

Keywords:

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10923558/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Host switching has contributed to the evolution of bacterial pathogens, yet the evolutionary processes enabling the emergence of host-associated lineages are not fully understood. Here, we provide new insights into the evolution of ST30 *Staphylococcus aureus* by characterizing animal-derived isolates in the context of human- and animal-derived genomes. Our findings are consistent with recurrent spillover of human-associated lineages into animals, whereas only camel-associated isolates formed a distinct monophyletic lineage characterized by extensive accessory genome remodeling. This remodeling involved gain and loss of mobile genetic elements, including loss of the immune evasion cluster and acquisition of a previously undescribed SCCcam carrying ACME type III, a staphylococcal pathogenicity island encoding an SdrD-like adhesin, a sea-carrying prophage, and a distinct accessory gene repertoire. Our findings suggest that recurrent spillover alone may be insufficient to establish stable host-associated populations and identify lineage-specific accessory genome remodeling as a prominent genomic feature accompanying sustained host-associated diversification following cross-species transmission.

Introduction

Cross-species transmission provides recurrent opportunities for bacterial pathogens to colonize new hosts; however, only a small fraction of these events culminates in the emergence of stable host-associated lineages. The evolutionary processes that determine whether host transitions remain transient spillover events or give rise to persistent host-associated populations remain poorly understood¹. In *S. aureus*, host shifts can be accompanied by extensive genomic remodeling, including the acquisition or loss of mobile genetic elements associated with immune evasion, virulence, and host specificity². Determining why some host-switching events result only in transient spillover while others culminate in stable host adaptation is therefore essential to understanding bacterial evolution.

Among *S. aureus* lineages, ST30 provides a valuable model for investigating these evolutionary processes. In the 1950s, the epidemic phage-type 80/81 (PT80/81) lineage, subsequently identified as methicillin-susceptible ST30 *S. aureus* (MSSA), caused widespread hospital outbreaks worldwide^{3,4}. Its global dissemination persisted until the early 1960s, when the introduction of methicillin reshaped the epidemiology of *S. aureus* infections^{4,5}.

Several decades later, in the early 2000s, another CC30 lineage, the community-acquired methicillin-resistant *S. aureus* (CA-MRSA) Southwest Pacific clone (SWP; also known as the Oceania Southwest Pacific clone, OSPC), emerged in Oceania and rapidly became prevalent^{6,7}. The clone subsequently spread to multiple continents, including South America and Europe, highlighting its remarkable epidemic potential^{8,9}. Both historical and contemporary epidemic ST30 lineages frequently harbor the *lukSF* genes encoding Panton–Valentine leukocidin (PVL)¹⁰, a cytotoxin associated with increased severity of *S. aureus* infections¹¹.

Beyond human hosts, *S. aureus* also colonizes and infects a wide range of animals, representing an important concern for both animal and public health. In livestock, the species is a major cause of diseases such as bovine mastitis, leading to substantial economic losses worldwide¹². In addition, livestock and companion animals can act as secondary reservoirs, providing opportunities for bidirectional transmission between humans and animals¹³.

Although the evolutionary history of human-associated ST30 lineages has been extensively investigated^{4,10,14,15}, the evolutionary origin of animal-associated CC30 remains poorly understood. Whether these isolates represent recurrent spillovers from human reservoirs or independently evolving host-associated lineages remains unknown. Here, we analyzed a global collection of 541 CC30 genomes from humans and animals to investigate the evolution of animal-derived ST30 isolates and identify the evolutionary processes underlying its emergence in animal hosts.

Results and discussion

Recurrent spillover predominates among animal-derived ST30 isolates

To investigate the evolutionary relationship of animal-derived ST30, we included all ST30 genomes available in the NCBI database for which metadata unambiguously identified an animal source. Phylogenomic reconstruction using 271 ST30 genomes (Supplementary Table 1, Sheet 1) revealed two distinct evolutionary patterns among animal-associated MSSA isolates. Eleven of the twelve camel isolates formed a previously unrecognized monophyletic lineage (G2-Sg1.2-Bn; brown in Fig. 1), whereas all remaining animal-derived isolates clustered within established human-associated ST30 clades previously defined by Esteves et al. (2025)¹⁰. These latter clades comprise isolates recovered from livestock, companion animals, wildlife, and poultry, with most (77.8%, 21/27) assigned to the human-associated subgroup G2-Sg2-Pr (purple in Fig. 1), followed by G2-Sg3-Gn (18.5%, 5/27) and G1-Pk (3.7%, 1/27) (green and pink; respectively in Fig. 1).

The phylogenetic placement of animal isolates embedded within human-associated clades is consistent with recurrent human-to-animal spillover rather than the emergence of stable host-adapted animal lineages. Instead of forming host-specific clusters, these animal-derived isolates remained interspersed among contemporary human-associated ST30 populations (Fig. 1). Although ST30 has previously been reported in livestock and companion animals^{16,17}, its global evolutionary dynamics across animal hosts have not previously been investigated.

One aspect of host adaptation in *S. aureus* is the loss of the immune evasion cluster (IEC), carried by a β -hemolysin-converting phage encoding human-specific immune evasion factors¹⁸. In livestock-associated ST398, for example, progressive IEC loss is considered a characteristic feature of adaptation to animal hosts¹⁹. However, this pattern was not consistently observed among animal-associated ST30

isolates. Most isolates embedded within human-associated clades retained IEC, including type A (40.7%, 11/27) and type B (18.5%, 5/27), whereas only 29.6% (8/27) lacked the element. Likewise, IEC loss was also detected among human-derived isolates (12.9%, 28/233) (Fig. 1).

The relatively higher frequency of IEC type A among animal isolates primarily reflected the phylogenetic composition of the dataset, as most belonged to the G2-Sg2-Pr lineage, in which this IEC-A predominates (Fig. 1)¹⁰. Conversely, the presence of IEC-negative animal isolates indicates that IEC distribution is not uniformly associated with host origin. Because IEC is phage encoded, its distribution is also expected to reflect recurrent phage gain and loss¹⁹, indicating that IEC absence alone should therefore not be interpreted as evidence of host adaptation.

The single animal-derived isolate assigned to G1-Pk (accession DAMQKV010000000), recovered in Europe in 1987, clustered within the historical phage-type 80/81 lineage responsible for the global epidemics of the 1950s and 1960s^{3,4}. Consistent with its phylogenetic position, the genome displayed the characteristic genomic features previously described for G1-Pk, including loss of most *lpl* genes within vSaa, vSaβ type III, IEC type B, SraP polymorphism A, and a SaPI-Tokyo-like pathogenicity island (Fig. 1; Table 1)¹⁰.

Table 1

Significant genomic events identified among 579 human- and animal-derived ST30 genomes, according to phylogenetic groups.

Genetic element	G1-Pk ^a	G2-Sg1.1-Yw	G2-Sg1.2-Bn	G2-Sg2-Pr	G2-Sg3-Gn
SaPlcam1	-	-	+ ^b	- ^b	-
SaPlcam2	-	-	+	-	-
SaPlcam3 ^c	-	-	+	-	-
vB_SauS_01_cam1 ^c	-	-	+	-	-
vB_SauS_01_cam2 ^c	-	-	+	-	-
SCCcam ^c	-	-	+	-	-
SCCmecIV	-	+	-	-	+
Toxin-antitoxin (toxM)	+	+	+	-	+
Lipoprotein-like complete cluster in _v_Saa ^d	-	+	+	+	+
ORFs near ESAT cluster ^e	-	+	+	+	+
φPVL108	+	-	-	-	-
Tn1421 carrying ssaA	-	-	-	+	-
Integrative plasmid carrying IS1272, and copB-mco	-	-	-	+	-
SaPI2 carrying tst	-	-	-	+	-
SaPI4	-	-	-	+	-
agrC/hla/isdH polymorphic alleles ^f	-	-	-	+	-
SaPI-14-021	-	-	-	-	+
pNTUH_6457	-	-	-	-	+
SraP-A ^g	+	-	-	-	-
SraP-B	-	+	-	-	-
SraP-C	-	-	-	+	-
SraP-D	-	-	-	-	+
SraP-E	-	-	+	-	-

Genetic element	G1-Pk ^a	G2-Sg1.1-Yw	G2-Sg1.2-Bn	G2-Sg2-Pr	G2-Sg3-Gn
SaPlcam1	-	-	+ ^b	- ^b	-
^a Phylogenetic groups and most associated significant events presented in this table were previously defined ¹⁰ . In the present study, the camel-associated lineage was identified within Group 2, leading to the subdivision of the previously defined G2-Sg1 into G2-Sg1.1 and G2-Sg1.2. Groups are designated as follows: G1-Pk (group 1, pink); G2-Sg1.1-Yw (group 2, subgroup 1.1, yellow); G2-Sg1.2-Bn (group 2, subgroup 1.2, brown); G2-Sg2-Pr (group 2, subgroup 2, purple); G2-Sg3-Gn (group 2, subgroup 3, green).					
^b +, detected in ≥ 80% of genomes in the respective group; -, detected in < 80%.					
^c Previously undescribed camel-specific mobile genetic elements identified in this study.					
^d Although some <i>lpl</i> genes were absent, the <i>lpl</i> cluster is more complete in G2 subgroups than in G1-Pk.					
^e While some differences were observed between G2-Sg2-Pr and the other subgroups, the identity among G2 genomes was substantially higher than that observed among those in G1-Pk.					
^f <i>agrC</i> encodes the accessory gene regulator C; <i>hla</i> encodes the hemolysin A; <i>isdH</i> , encodes the iron-regulated surface determinant protein H.					
^g SraP-A to SraP-E represent the different SraP variants identified among ST30 lineages.					

Similarly, animal isolates assigned to Group 2 retained the genomic characteristics of their respective phylogenetic subgroups (Supplementary Table 1, Sheets 2 and 3).

Isolates within G2-Sg2-Pr exhibited the characteristic *tst/sea*-positive MSSA profile, whereas those assigned to G2-Sg3-Gn displayed the PVL-positive genotype (Fig. 1; Table 1). The conservation of additional subgroup-specific features, including *agr* polymorphisms, SraP classification, and mobile genetic elements, further supports recurrent human-to-animal spillovers rather than the emergence of independently evolving animal-adapted ST30 lineages (Fig. 2; Table 1). Thus, most animal-derived ST30 isolates retained the genomic signatures of their corresponding human-associated sublineages, rather than forming distinct animal-associated population. In contrast, the camel-associated isolates formed a genetically distinct monophyletic lineage and were therefore investigated in greater detail (Fig. 1; Table 1 and Supplementary Table 1, Sheets 2 and 3).

A distinct camel-associated ST30 lineage underwent extensive accessory genome remodeling

In contrast to the recurrent spillover pattern observed among the remaining animal-derived ST30 isolates, only a single camel-derived isolate clustered within the human-associated G2-Sg3-Gn lineage

(Fig. 1), whereas the remaining eleven formed the G2-Sg1.2-Bn lineage, which was phylogenetically distinct within the available global ST30 genomic diversity. None of the camel-associated lineage/subgroup G2-Sg1.2-Bn carried the IEC, compared with 205/233 (87.9%) human-derived isolates and 19/27 (70.3%) animal isolates embedded within human-associated phylogenetic groups ($p < 0.0001$ and $p = 0.00013$; respectively) (Supplementary Table 1, Sheets 2–4). The complete absence of IEC distinguishes the camel-associated subgroup from all other ST30 clusters analyzed and is consistent with the loss of human-specific immune evasion functions that was observed during host transitions in *S. aureus*¹⁸.

In agreement with its distinct phylogenetic position, the camel-associated subgroup carried a unique SraP polymorphism that was absent from all other human and animal ST30 genomes analyzed^{10,15}. This previously undescribed variant was designated SraP-E (Fig. 2).

Bayesian time-scaled phylogenetic analysis confirmed that most camel-associated isolates formed a distinct monophyletic lineage. The most recent common ancestor (tMRCA) of this lineage was estimated to have existed around 1920 (95% HPD: 1874–1955). Although the 95% HPD interval is broad, the estimated age argues against a very recent origin of this lineage. Internal nodes within the camel-associated clade were dated predominantly between 1995 and 2005, suggesting more recent diversification following its establishment. The camel-associated isolates formed a cohesive cluster with no human isolates interspersed within this subgroup, a phylogenetic pattern consistent with sustained diversification of a single camel-associated lineage rather than multiple independent human-to-camel transmission events (Extended Data Fig. 1).

Despite their extensive genomic divergence, camel-associated ST30 isolates retained a virulence gene repertoire largely comparable to that of other CC30 populations, including the major adhesins (*clfA*, *clfB* and *cna*) and *hla*. However, they lacked the *lukS-PV/lukF-PV* genes encoding PVL (Supplementary Table 1, Sheets 2–4), a toxin frequently associated with epidemic ST30 lineages^{10,15}. Camel-associated genomes retained the genomic islands vSa α and vSa β type III, which are widely distributed among ST30 lineages¹⁰. In addition, all isolates from the camel-associated lineage carried the previously described staphylococcal pathogenicity island SaPIcam1²⁰, although this element does not encode any known virulence or antimicrobial resistance determinants.

Notably, the *tst* gene encoding toxic shock syndrome toxin-1 (TSST-1) was present in approximately half of the camel-associated isolates and was also common among animal isolates embedded within human-associated phylogenetic groups (Fig. 1). However, unlike the canonical ST30 population, in which *tst* is typically carried by SaPI2¹⁰, camel-associated lineage harbored *tst* within SaPIcam2, a pathogenicity island previously identified in camel-derived *S. aureus*²⁰. Although SaPIcam2 was originally reported to carry a candidate superantigen *tst* gene²⁰, comparative sequence analysis performed in the present study found that the encoded protein is identical to the TSST-1 of the prototype menstrual toxic shock syndrome strain MN8 (NCBI protein accession: WP_001035597.1), exhibiting 100% amino acid identity across the full-length sequence.

Here we identified a previously undescribed SaPI-like element detected exclusively in the camel-associated ST30 subgroup analyzed here. Whole-sequence BLAST analysis did not identify any previously described SaPI showing full-length identity to this element, indicating that it represents a distinct SaPI-like pathogenicity island, restricted to the camel-associated ST30 lineage, here designated SaPIcam3 (Supplementary Table 2, Sheet 1; Extended Data Fig. 2). Comparative genomic analysis indicated that the integrative element spans approximately 18.8 kb in the reference genome (NCBI accession: GCF_025559685.1), corresponding to positions 443,621–462,403. However, because the canonical *attL* and *attR* recombination sites could not be unambiguously identified by computational analysis, the boundaries of the SaPIcam3 remain putative and were inferred from comparative genomic alignment with ST30 genomes lacking the island (Fig. 3). This island comprises 29 open reading frames (ORFs), including genes involved in integration, regulation (*xre*) family, replication/maintenance, packaging/mobilization, several SaPI-associated hypothetical ORFs, and a single accessory cargo gene encoding a putative Sdr-family protein, here designated *sdrD-like* (Fig. 3; Supplementary Table 3). Members of the Sdr family belong to the microbial surface components recognizing adhesive matrix molecules (MSCRAMMs), a group of surface adhesins involved in bacterial attachment to host extracellular matrix proteins, including fibrinogen and fibronectin²¹.

Remarkably, the *sdrD-like* gene was absent from all 541 human-derived ST30 genomes analyzed and from all 27 animal isolates embedded within the human-associated ST30 lineages. BLAST searches identified homologous *sdr-like* sequences exclusively in animal-associated *S. aureus* isolates from northern Kenya, including camel and canine strains (Supplementary Table 2, Sheet 2). Among the camel isolates, SdrD-like homologs were detected in multiple sequence types, whereas all canine isolates belonged to ST1292, indicating that this adhesin is not confined to a single *S. aureus* lineage. Although the genomic context of these homologs remains to be determined, their distribution is consistent with horizontal interclonal dissemination of the SdrD-like-encoding gene, or of a larger genetic module carrying it, rather than exclusive vertical inheritance within a single lineage (Extended Data Fig. 3; Supplementary Table 2, Sheet 2).

The discordance between the restricted distribution of the complete SaPIcam3 and the broader occurrence of the *sdrD-like* gene suggests that this locus has an evolutionary history partly independent of the complete island. SaPIcam3 contains multiple recombination-associated features, including integrase/recombinase-related genes and several conserved SaPI-associated ORFs, providing a potential genetic basis for processes such as intragenomic rearrangements, partial island mobilization, or recombination with other SaPI-like elements²².

Although these mechanisms remain to be experimentally demonstrated, they provide a plausible explanation for the dissemination of the *sdrD-like* gene in the apparent absence of the complete SaPIcam3 backbone. In ecosystems where multiple host species coexist²³, such ecological interactions may facilitate bacterial exchange and the dissemination of mobile genetic elements or their associated genes, highlighting the potential One Health relevance of accessory gene dissemination across animal-associated bacterial populations.

Consistent with its annotation as an SdrD-family protein (Supplementary Table 3), comparative structural modeling, supported by sequence alignment of the conserved Sdr-family ligand-binding region, demonstrated that the SdrD-like sequence retained a region structurally homologous to the ligand-binding region of canonical SdrD²⁴, whereas the LPXTG sorting motif and the serine-aspartate (SD) repeat region were absent. Structural comparison further revealed that this conserved region represents a truncated version of the canonical ligand-binding architecture, with partial loss of the region corresponding to the canonical N2 domain (Extended Data Fig. 4).

Although the LPXTG motif is typically required for sortase-mediated anchoring of MSCRAMM proteins to the peptidoglycan, its absence does not necessarily preclude biological activity. For example, the *S. aureus* immune evasion protein Sbi lacks an LPXTG motif yet remains associated with the bacterial surface through non-covalent interactions while retaining IgG-binding activity²⁵. The contribution of the SdrD-like protein to camel colonization or mastitis remains unknown; however, SdrD homologues have been reported in bovine and ovine mastitis-associated isolates, where they contribute to adhesion and persistence within mammary tissue²⁶.

The identification of a previously unrecognized SdrD-like sequence prompted us to investigate its evolutionary history. Codon-based analysis using aBSREL detected episodic diversifying selection on multiple branches across phylogeny. However, no evidence of episodic diversifying selection was detected along the branch defining the camel-associated SdrD-like lineage, its immediate ancestral branch, or camel terminal branch (Supplementary Fig. 1; Supplementary Table 4, Sheets 1–5), indicating that episodic diversifying selection was not specifically associated with the camel lineage. The distribution of the *sdrD-like* gene across the camel-associated clade, together with its conserved genomic association with SaPlcam3, is consistent with acquisition before diversification of this clade, likely through horizontal transfer of SaPlcam3, in agreement with the well-established capacity of SaPIs for mobilization²².

Unlike most of the ST30 population, in which *sea* is typically carried by the IEC¹⁰, all isolates from camel-associated lineage harbored a previously undescribed temperate bacteriophage carrying a *sea-like* gene. Although *sea-like* was detected in most camel-associated isolates and canonical *sea* was common among animal isolates embedded within human-associated phylogenetic groups, only the camel-associated subgroup consistently carried *sea-like* within this prophage (Fig. 1, Fig. 4; Supplementary Table 5, Sheets 1 and 2).

An insertion of the IS1182 element into the 5' region of *sea* disrupted the canonical gene architecture, removed the 5' regulatory/start region, and the N-terminal coding sequence. Consequently, the original *sea* ORF was replaced by a shorter downstream *sea-like* ORF lacking the first ~ 41 amino acids of the curated reference SEA protein (UniProt accession number: P0A0L2) (Extended Data Fig. 5). Whether this ORF is transcriptionally silent or instead transcribed from regulatory sequences provided or activated by the adjacent IS1182 element remains unknown. Experimental studies will be required to determine whether this remodeled *sea-like* ORF is expressed and retains biological activity. The phage exhibited the

characteristic genomic organization of the *Siphoviridae* morphotype²⁷, including genes encoding an integrase, terminase, tape measure protein and long-tail structural proteins, and spanned approximately 39.9 kb. The prophage was integrated at positions 1,657,975–1,697,892 of the reference genome (GCF_025559685.1) and was designated vB_SauS_01_cam1. Despite also carrying a *sea-like* gene, PHASTEST and BLAST analysis indicated that vB_SauS_01_cam1 was genomically distinct from sea-carrying ϕ Sa3 prophages previously reported in non-ST30 *S. aureus* (Extended Data Fig. 5; Supplementary Table 5, Sheet 2).

In addition, we identified another previously undescribed prophage that lacks recognizable virulence and antimicrobial resistance genes. PHASTEST classified the element as an intact 44.8-kb prophage comprising 59 predicted ORFs (Supplementary Table 5, Sheets 3 and 4). Comparative analysis revealed a mosaic genome architecture, with the highest similarity to *Staphylococcus* phage vB_SauS_phi2; however, only 25 of the 59 predicted proteins (42.4%) identified this phage as their closest homolog, whereas the remaining proteins showed highest similarity to multiple distinct staphylococcal phages, consistent with the modular evolution of temperate phages²⁷. The prophage spans positions 349,050–393,946 in the reference genome (NCBI accession no. GCF_025559685.1) and was designated vB_SauS_01_cam2. (Extended Data Fig. 6).

We also identified a previously undescribed non-*mec* staphylococcal cassette chromosome (SCC) integrated adjacent to *rlmH* (*orfX*) in the reference genome (NCBI accession no. GCF_025559685.1) (Supplementary Table 6, Sheet 1; Extended Data Fig. 7). Comparative sequence analysis identified the structural hallmarks of SCC elements²⁸, including cassette chromosome recombinase (*ccr*) genes, *attL* and *attR* attachment sites delimiting the element, and the absence of the *mec* complex. Based on these structural features, the element was designated SCCcam. SCCcam exhibits a highly mosaic architecture composed of multiple functional modules acquired through successive recombination events. In addition to an arsenic resistance (*arsRBC*) operon, which mediates arsenic detoxification through the reduction of arsenate and active efflux of arsenite from the bacterial cell²⁹, the element harbors an ACME type III module displaying high sequence similarity and conserved synteny with ACME III elements identified in diverse bacterial species, including *Staphylococcus epidermidis*³⁰ (Fig. 5; Supplementary Table 6, Sheet 2).

Beyond the ACME III region, SCCcam comprises a diverse repertoire of functional modules associated with environmental adaptation, genome maintenance, and genome plasticity. These include recombination functions encoded by *ccrA1/ccrB1*, heavy metal resistance determinants (*arsRBC* and *chrA*), stress adaptation genes encoding universal stress protein (UspA), sulfide oxidoreductase and rhodanese proteins, an OPP oligopeptide permease operon constituting the ACME III metabolic module, DNA maintenance functions including *rloC*, *nadD*, and a remnant *radC* gene, cell wall remodeling enzymes (LytM endopeptidases), DNA methylation functions, an IS30-family transposase, and several conserved hypothetical proteins. The combination of these functionally distinct modules supports the conclusion that SCCcam represents a previously undescribed composite SCC-like element assembled

through sequential acquisition and remodeling of mobile DNA segments (Fig. 5; Supplementary Table 6, Sheet 2).

Although an SCC-associated ACME III has previously been reported in *S. aureus* ST80³¹, BLASTn comparison of the complete SCC regions demonstrated that SCCcam shares extensive homology with previously described SCC-associated ACME elements only across the ACME III module, whereas the remaining SCC backbone exhibits little or no detectable similarity. These findings indicate that SCCcam is not a variant of previously described SCC elements but rather represents a distinct composite SCC-like element that independently acquired an ACME III module (Extended Data Fig. 8).

The identification of SCCcam, together with the previously undescribed prophages and the unique pathogenicity islands SaPlcam3, in addition to the previously described SaPlcam1 and SaPlcam2²⁰, reveals that the emergence of the camel-associated ST30 lineage was accompanied by extensive remodeling of its accessory genome through the acquisition, diversification, and retention of multiple mobile genetic elements. Together, these findings indicate that horizontal gene transfer has been a major evolutionary force shaping the genomic architecture of the camel-associated lineage, although the contribution of these accessory elements to host adaptation remains to be established.

Several limitations should be considered when interpreting our findings. The number of publicly available animal-derived ST30 genomes remains limited, particularly MSSA isolates, and no human-derived ST30 genomes from the geographic regions where the camel isolates were collected were available for comparison. Consequently, although our analyses support the existence of a genetically distinct camel-associated lineage characterized by multiple lineage-specific mobile genetic elements, they cannot definitively distinguish long-term host adaptation from an unsampled regional human reservoir. Notably, however, the only camel isolate that clustered within the human-associated cluster was assigned to G2-Sg3-Gn, despite its geographic separation from the isolates comprising this subgroup. In addition, the biological role of the SdrD-like variant and functional significance of the sea-like gene remain to be experimentally validated.

Overall, animal-associated ST30 isolates reflect distinct evolutionary trajectories, ranging from recurrent spillover events involving human-associated lineages to sustained host-associated diversification following cross-species transmission. The findings indicate that recurrent spillover of human-associated lineages into animals was the predominant pattern observed among animal-derived ST30 isolates. In contrast, the camel-associated lineage followed a distinct evolutionary trajectory, characterized by long-term diversification and the acquisition of lineage-specific mobile genetic elements. From a One Health perspective, repeated establishment of human-associated ST30 lineages in livestock, companion animals and wildlife may contribute to shared *S. aureus* reservoirs, creating opportunities for bidirectional transmission and the emergence of host-adapted variants¹³. In livestock, particularly dairy cattle, such spillover events may also have important economic consequences through their contribution to persistent intramammary infections and bovine mastitis^{12,26}. Our findings broaden the current

understanding of ST30 evolution beyond the human host and underscore the importance of integrating animal reservoirs into global genomic surveillance.

Methods

Genome selection

A curated dataset comprising 541 human-derived ST30 genomes, previously analyzed in our study of global ST30 evolution¹⁰, was used as the reference collection. Most genomes were retrieved from the NCBI database and represented multiple geographic regions worldwide, whereas 37 Brazilian genomes were sequenced by us. For phylogenomic analyses, 233 human ST30 genomes were selected by proportionally stratified random sampling to represent the major phylogenetic groups and subgroups previously defined¹⁰. In addition, all publicly available animal-associated ST30 genomes with relevant metadata ($n = 35$) and three bovine ST30 genomes sequenced in this study were included, yielding a final dataset of 271 genomes (Supplementary Table 1, Sheet 1).

Whole-genome sequencing, annotation and phylogenetic reconstruction

Genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Whole-genome sequencing was performed by MicrobesNG (Birmingham, UK) on an Illumina platform generating 2×250 bp paired-end reads. Raw-read quality control, genome assembly, and annotation were performed using the provider's standard bioinformatics pipeline.

Phylogenomic analyses were conducted using a dataset of 271 *S. aureus* ST30 genomes from human and animal sources. The ST239 (CC8)-SCCmecIII strain BMB9393 (GCF_000418345.1) and the ST5 (CC5)-SCCmecII strain N315 (GCF_000009645.1) were included as outgroups. Genomes were annotated with Prokka v1.14.6³². In addition, functional annotation of hypothetical proteins was manually refined by combining sequence similarity searches (BLASTP and UniProt), conserved domain prediction (InterPro and CDD), and remote homology detection using HHpred. Core-genome alignments were generated using Roary v3.13.0 with MAFFT, retaining genes present in $\geq 99\%$ of the genomes³³. Maximum-likelihood phylogenies were reconstructed with RAxML v8.2.12 under the GTR-CAT substitution model with 1,000 bootstrap replicates³⁴, rooted using the outgroup genomes, and visualized in iTOL v7.4.1³⁵.

Virulence, resistance, and mobile genetic elements

Virulence-associated genes in the 38 animal-associated isolates were identified by BLAST +³⁶ against the Virulence Factor Database (VFDB)³⁷, using minimum thresholds of 90% sequence coverage and 95% nucleotide identity. Virulence profiles were subsequently compared with those of the 541 human-derived

genomes. To avoid misannotation of paralogous leukocidin genes, *lukS-PV* and *lukF-PV* were queried as a single contiguous genomic region from *S. aureus* strain 13420 (NZ_CP021141.1; positions 1,518,664–1,520,582).

Genomic islands (GIs), including vSaa, vSaβ, vSay, SaPI1, SaPI2, SaPI3, SaPI4, and immune evasion cluster (IEC) types A–F, were identified using previously described genetic markers, and their genomic organization was assessed as described previously¹⁰. ORFs corresponding to each GI and SaPI were extracted in Geneious Prime v2026.0.2 (Biomatters, Auckland, New Zealand) using the Map to Reference function and queried against the complete dataset (541 human and 38 animal genomes) by BLAST, applying thresholds of 95% nucleotide identity and 90% coverage. Acquired antimicrobial resistance genes were identified using ABRicate with the Comprehensive Antibiotic Resistance Database (CARD) and a minimum nucleotide identity of 90%. Prophages were identified with PHASTEST³⁸, and phage regions were extracted in Geneious Prime and manually annotated using UniProt and NCBI.

Comparisons of IEC frequencies between the camel-associated lineage and the other groups were performed using Fisher's exact test.

ORF-based pangenomic analysis

An open reading frame (ORF)-based pangenomic analysis was performed as previously described¹⁰ to investigate ORF variation and the distribution of mobile genetic elements (MGEs). Pangenome groups were defined using genomes from both human and animal hosts. Reference genomes were selected for each group: strain IVB6185 (GCA_025559605.1) for the animal-associated dataset and strain CR14-021 (GCA_021012415.1) for the human-associated dataset. ORFs were compared across genomes using command-line BLAST searches, and hits with < 90% query coverage were excluded. For hits meeting the ≥ 90% coverage criterion, sequence identity values ≥ 90% were assigned a value of 1, and those < 90% were assigned a value of 0. Differences in ORF distribution between animal- and human-associated genomes were assessed for each ORF using two-sided Fisher's exact tests, with *p* values adjusted using Benjamini–Hochberg false discovery rate (FDR) correction. ORFs with an FDR-adjusted *p* value < 0.05 were considered statistically significant and selected for downstream analysis. In addition, sequence identity patterns were visually inspected. To facilitate the identification of group-specific patterns and contiguous ORF blocks differing between groups, genome names were color-coded according to the groups defined in the phylogenetic trees, using the same color scheme, with the camel-associated lineage visually separated from the remaining lineages, while sequence identity values < 90% and ≥ 90% were displayed using distinct colors. ORFs were mapped using the Map to Reference tool in Geneious Prime v2026.02 (Biomatters, Auckland, New Zealand).

Lineage associated pangenome analysis

To further identify camel-lineage-associated accessory genes rather than infer causal determinants of camel adaptation, this analysis was performed using 22 ST30 genomes, comprising all 11 camel-associated isolates and 11 human isolates selected from the phylogenetically closest clades. Selection of closely related human isolates minimized confounding due to population structure while preserving the power to detect lineage-specific accessory genes. Gene presence–absence matrices generated from the pangenome analysis were used for association testing. Gene–lineage associations were initially evaluated using Scoary v1.6.16³⁹, with lineage membership (camel-associated versus human-associated) defined as the binary phenotype. Statistical significance was assessed using Fisher's exact test followed by Benjamini–Hochberg false discovery rate (FDR) correction.

Bayesian phylogenetic reconstruction and temporal analysis

Bayesian time-scaled phylogenetic analyses were performed using a representative dataset of 74 genomes, comprising all 11 camel-associated isolates and 63 human isolates selected to capture the phylogenetic diversity of the global CC30 population. Core-genome alignments generated with Roary v3.13.0 were screened for recombination using ClonalFrameML v1.13, and recombinant regions were excluded prior to molecular clock analyses. Bayesian inference was performed in BEAST v1.10.4⁴⁰ under the HKY substitution model with estimated base frequencies, gamma-distributed rate heterogeneity, an uncorrelated log-normal relaxed molecular clock, and a constant-size coalescent prior. The molecular clock prior was centered on a substitution rate of 1.5×10^{-6} substitutions per site per year. Two independent Markov chain Monte Carlo (MCMC) chains were run for 200 million generations, with the first 10% discarded as burn-in. Convergence was assessed in Tracer v1.7.1, and all parameters achieved effective sample sizes (ESS) > 200. Maximum clade credibility trees were generated with TreeAnnotator v1.10.4 and visualized in iTOL v7. The mean substitution rate estimated by BEAST was 3.5×10^{-6} substitutions per site per year (95% HPD: $1.11 \times 10^{-6} - 4.11 \times 10^{-6}$), consistent with previous estimates for *S. aureus*⁴¹.

Structural modeling and domain comparison of Sdr proteins

Putative Sdr proteins were modeled using AlphaFold2 through the ColabFold implementation⁴² with default parameters. For each sequence, the model with the highest predicted local distance difference test (pLDDT) score was selected. Predicted structures were compared with experimentally characterized *S. aureus* SdrD proteins using UCSF ChimeraX⁴³. Structural similarity was assessed by pairwise superposition with the MatchMaker algorithm, root mean square deviation (RMSD), and comparison of secondary-structure topology and domain organization. Conserved amino acid residues identified by multiple sequence alignment were mapped onto the predicted structures, with particular emphasis on the N2 and N3 subdomains of the A region involved in the dock–lock–latch (DLL) ligand-binding mechanism of MSCRAMM proteins²⁴.

For domain annotation, protein sequences were analyzed with HMMER v3.4⁴⁴ against the Pfam-A database (release 37.0) using hmmscan with default parameters, and independently annotated with InterProScan v5⁴⁵, integrating the Pfam, TIGRFAM, CDD, SUPERFAMILY, and PANTHER databases. Domain architectures, including signal peptides, ligand-binding domains, B-repeat regions, and LPXTG cell wall-anchoring motifs, were compared between canonical human-associated Sdr proteins and animal-associated homologs.

Phylogenetic analysis of Sdr proteins and estimation of selective pressure (dN/dS)

Representative Sdr proteins from the major *S. aureus* clonal complexes were included in the phylogenetic analysis, comprising five sequences each from CC5, CC8, CC22, CC30, and CC398 ($n = 25$). Five additional animal-associated homologs, including four camel-associated and one dog-associated sequence identified in publicly available genomes, were incorporated. Two representative SdrC proteins (NCTC8325 and USA300) and two SdrE proteins (COL and Newman) were included as outgroups, yielding a final dataset of 34 Sdr-related proteins. Protein sequences were retrieved from UniProt and aligned using MACSE v2.07⁴⁶. Ambiguously aligned regions were removed with trimAl v1.4.1 using the automated1 setting⁴⁷, followed by manual inspection to preserve codon integrity. Maximum-likelihood phylogenies were reconstructed with IQ-TREE2 v2.2.0⁴⁸. The best-fitting amino acid substitution model was selected using ModelFinder according to the Bayesian Information Criterion (BIC), and branch support was estimated with 1,000 ultrafast bootstrap replicates. Codon-aware alignments generated with MACSE were analyzed using the adaptive Branch-Site Random Effects Likelihood (aBSREL) model implemented in the Datamonkey server^{49,50} to identify branches under episodic diversifying selection. Statistical significance was assessed using likelihood-ratio tests with Holm–Bonferroni correction for multiple testing.

Use of artificial intelligence

During manuscript preparation, ChatGPT (OpenAI) was used to assist with English-language editing and to suggest improvements to the visual design of selected figures. All figures were subsequently prepared by the authors using the original study data, and no AI-generated images or artwork were included in the manuscript. All scientific content, data interpretation and visual presentation were reviewed and verified by the authors, who take full responsibility for the final manuscript.

Reporting summary

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

Data availability

Whole-genome sequencing data have been deposited in the NCBI database, and the corresponding accession numbers are listed in Supplementary Table 1 (Sheet 1). All datasets generated and analyzed

during this study, including genome annotations, BLAST analyses, Roary outputs, pangenome matrices, and other supporting datasets, are available in the Supplementary Tables. Source data are provided with this paper.

Declarations

Author contributions

M.A.C.E.: Investigation, Formal analysis, Methodology, Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing.

M.F.C.: Investigation, Formal analysis, Methodology, Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing.

L.M.A.O.: Investigation, Methodology, Formal analysis, Data curation, Writing – review & editing.

C.L.M.: Investigation, Methodology, Formal analysis, Data curation, Supervision, Writing – review & editing.

L.G.A.L.: Investigation, Methodology, Formal analysis, Data curation, Supervision, Writing – review & editing.

T.C.A.P.: Investigation, Methodology, Formal analysis, Data curation, Supervision, Writing – review & editing.

A.M.S.F.: Project administration, Conceptualization, Methodology, Formal analysis, Data curation, Validation, Visualization, Supervision, Resources, Funding acquisition, Writing – review & editing.

Funding

This work was partly supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grant numbers 306942/2023-2; and 408677/2024-4, INCT-SOMAR), the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (grant number E-26/203.941/2024 and E-26/210.806/2025) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; grant number 001).

Competing interests

The authors declare no competing interests.

References

1. Sheppard, S. K., Guttman, D. S., & Fitzgerald, J. R. Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* 19, 549–565 (2018).

2. Laumay, F., Benchetrit, H., Corvaglia, A. R., van der Mee-Marquet, N. & François, P. The *Staphylococcus aureus* CC398 lineage: an evolution driven by the acquisition of prophages and other mobile genetic elements. *Genes* 12, 1752 (2021)
3. Shaffer, T. E., Sylvester, R. F., Baldwin, J. N. & Rheins, M. S. Staphylococcal infections in newborn infants. II. Report of 19 epidemics caused by an identical strain of *Staphylococcus pyogenes*. *Am. J. Public Health Nations Health* 47, 990–994 (1957).
4. DeLeo, F. R. et al. Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* 108, 18091–18096 (2011).
5. Jevons, M. P. 'Celbenin'-resistant staphylococci. *Br. Med. J.* 1, 124–125 (1961).
6. Huygens, F., Nimmo, G. R., Schooneveldt, J., Munckhof, W. J. & Giffard, P. M. Genotyping of methicillin-resistant *Staphylococcus aureus* by assaying for the presence of variable elements associated with *mecA*. *J. Clin. Microbiol.* 40, 3093–3097 (2002).
7. Okuma, K. et al. Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. *J. Clin. Microbiol.* 40, 4289–4294 (2002).
8. Ribeiro, A. et al. Detection and characterization of international community-acquired infections by methicillin-resistant *Staphylococcus aureus* clones in Rio de Janeiro and Porto Alegre cities causing both community- and hospital-associated diseases. *Diagn. Microbiol. Infect. Dis.* 59, 339–345 (2007).
9. Rossney, A. S. et al. The emergence and importation of diverse genotypes of methicillin-resistant *Staphylococcus aureus* (MRSA) harboring the Panton–Valentine leukocidin gene (*pvl*) reveal that *pvl* is a poor marker for community-acquired MRSA strains in Ireland. *J. Clin. Microbiol.* 45, 2554–2563 (2007).
10. Esteves, M. A. C. et al. Decoding the evolutionary history of ST30 *Staphylococcus aureus*: insights into a potentially silent MSSA bloodstream pathogen. *Frontiers in microbiology*, 16, 1522747 (2025).
11. Crémieux, A.-C. et al. Panton-Valentine leukocidin enhances the severity of community-associated methicillin-resistant *Staphylococcus aureus* rabbit osteomyelitis. *PLoS One* 4, e7204 (2009).
12. Fadillah, A., Van Den Borne, B. H. P., Schukken, Y. H., Poetri, O. N. & Hogeveen, H. Cost-efficiency of mastitis control strategies on smallholder dairy farms. *J. Dairy Sci.* 108, 8726–8741 (2025).
13. Machado da Silva, I. N. et al. Environmental and animal reservoirs of methicillin-resistant *Staphylococcus aureus* (MRSA): a One Health challenge. *J. Toxicol. Environ. Health B Crit. Rev.* 29, 329–357 (2026).
14. McGavin, M. J., Arsic, B. & Nickerson, N. N. Evolutionary blueprint for host- and niche-adaptation in *Staphylococcus aureus* clonal complex CC30. *Front. Cell. Infect. Microbiol.* 2, 48 (2012).
15. Di Gregorio, S. et al. Genomic epidemiology of CC30 methicillin-resistant *Staphylococcus aureus* strains from Argentina reveals four major clades with distinctive genetic features. *mSphere* 6, e01297-20 (2021).
16. Lozano, C., Gharsa, H., Ben Slama, K., Zarazaga, M. & Torres, C. *Staphylococcus aureus* in animals and food: methicillin resistance, prevalence and population structure. A review in the African

- continent. *Microorganisms* 4, 12 (2016).
17. Mohamad Farook, N. A. et al. Diversity and dissemination of methicillin-resistant *Staphylococcus aureus* (MRSA) genotypes in Southeast Asia. *Trop. Med. Infect. Dis.* 7, 438 (2022).
 18. Bouiller, K., Bertrand, X., Hocquet, D. & Chirouze, C. Human infection of methicillin-susceptible *Staphylococcus aureus* CC398: a review. *Microorganisms* 8, 1737 (2020).
 19. Rohmer, C. & Wolz, C. The role of hlb-converting bacteriophages in *Staphylococcus aureus* host adaption. *Microb. Physiol.* 31, 109–122 (2021).
 20. Zubair, S. et al. Complete genome sequence of *Staphylococcus aureus* strain ILRI_Eymole1/1, isolated from a Kenyan dromedary camel. *Stand. Genomic Sci.* 10, 109 (2015).
 21. 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).
 22. Novick, R. P. Pathogenicity islands and their role in staphylococcal biology. *Microbiol. Spectr.* 7, GPP3-0062-2019 (2019).
 23. Watson, E. E., Kochore, H. H. & Dabasso, B. H. Camels and climate resilience: adaptation in northern Kenya. *Hum. Ecol.* 44, 701–713 (2016).
 24. Ponnuraj, K. et al. A ‘dock, lock, and latch’ structural model for a staphylococcal adhesin binding to fibrinogen. *Cell* 115, 217–228 (2003).
 25. Burman, J. D. et al. Interaction of human complement with Sbi, a staphylococcal immunoglobulin-binding protein: indications of a novel mechanism of complement evasion by *Staphylococcus aureus*. *J. Biol. Chem.* 283, 17579–17593 (2008).
 26. Campos, B. et al. Diversity and pathogenesis of *Staphylococcus aureus* from bovine mastitis: current understanding and future perspectives. *BMC Vet. Res.* 18, 115 (2022).
 27. Goulet, A., Spinelli, S., Mahony, J. & Cambillau, C. Conserved and diverse traits of adhesion devices from *Siphoviridae* recognizing proteinaceous or saccharidic receptors. *Viruses* 12, 512 (2020).
 28. Liu, J. et al. Staphylococcal chromosomal cassettes *mec* (SCC*mec*): a mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microb. Pathog.* 101, 56–67 (2016).
 29. Yang, H.-C. & Rosen, B. P. New mechanisms of bacterial arsenic resistance. *Biomed. J.* 39, 5–13 (2016).
 30. McManus, B. A. et al. First detailed genetic characterization of the structural organization of type III arginine catabolic mobile elements harbored by *Staphylococcus epidermidis* by using whole-genome sequencing. *Antimicrob. Agents Chemother.* 61, e01216-17 (2017).
 31. Gawlik, D. et al. Molecular investigations on a chimeric strain of *Staphylococcus aureus* sequence type 80. *PLoS One* 15, e0232071 (2020).
 32. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069 (2014).
 33. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013).

34. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).
35. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296 (2021).
36. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421 (2009).
37. Zhou, Siyu, Bo Liu, Dandan Zheng, Lihong Chen, Jian Yang. “VFDB 2025: An integrated resource for exploring anti-virulence compounds”. *Nucleic Acids Res.* 53: D871–D877 (2025).
38. Arndt, D., Marcu, A., Liang, Y. & Wishart, D. S. PHAST, PHASTER and PHASTEST: tools for finding prophage in bacterial genomes. *Brief. Bioinform.* 20, 1560–1567 (2019).
39. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17, 238 (2016).
40. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4, vey016 (2018).
41. Duchêne, S. et al. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* 2, e000094 (2016).
42. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682 (2022).
43. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82 (2021).
44. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37 (2011).
45. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240 (2014).
46. Ranwez, V., Chantret, N. & Delsuc, F. Aligning protein-coding nucleotide sequences with MACSE. *Methods Mol. Biol.* 2231, 51–70 (2021).
47. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009).
48. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534 (2020).
49. Smith, M. D. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353 (2015).
50. Weaver, S. et al. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777 (2018).

Figures

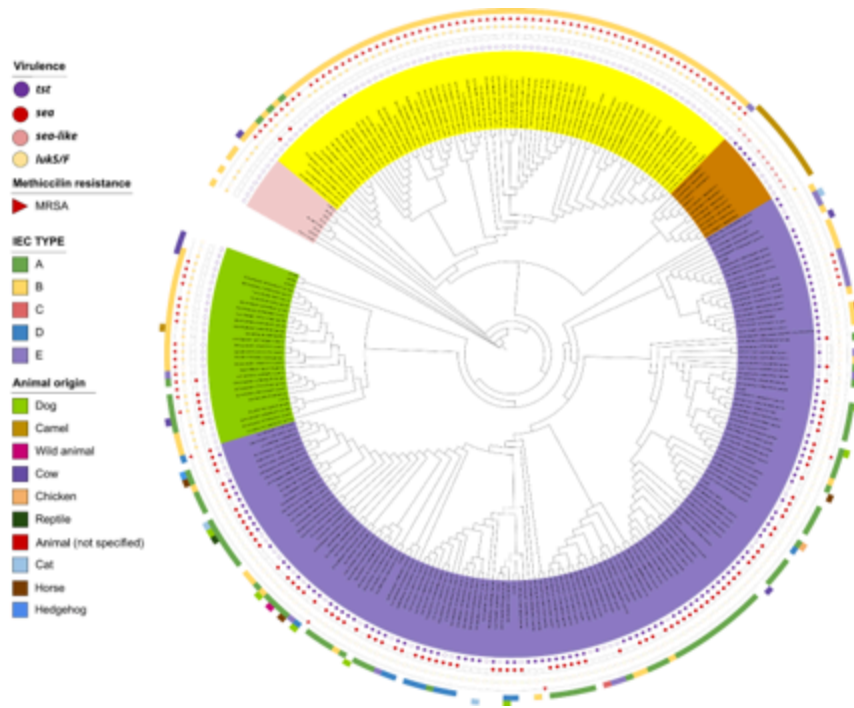


Figure 1

Global phylogenetic structure of animal- and human-derived *S. aureus* ST30 isolates. Maximum-likelihood phylogeny reconstructed from 271 ST30 genomes of human- and animal- derived genomes. Phylogenetic groups are indicated by background colors: G1-Pk (pink), G2-Sg1.1-Yw (yellow), G2-Sg1.2-Bn (brown), G2-Sg2-Pr (purple) and G2-Sg3-Gn (green). Virulence-associated genes, including *tst*, *sea*, *sea-like* and *lukS/F* genes, encoding toxic shock syndrome toxin-1, enterotoxin A, enterotoxin A-like protein, and Panton Valentine leukocidin, respectively, are indicated by colored circles. Methicillin-resistant *S. aureus* (MRSA) is indicated by a red triangle. Immune evasion cluster (IEC) types are represented by colored squares in the inner annotation ring. Animal-derived isolates are indicated by colored squares in the outermost annotation ring according to host origin.

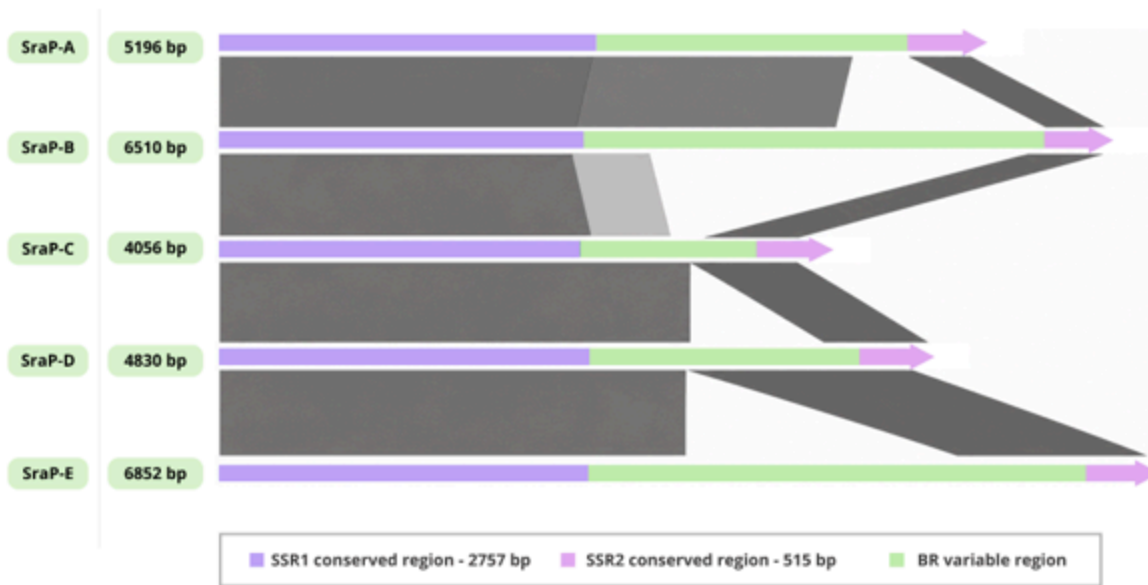


Figure 2

Structural polymorphism of *sraP* among ST30 isolates. Schematic representation of five *sraP* polymorphic patterns (SraP-A to SraP-E) identified among the different phylogenetic groups. Each pattern is shown according to its predicted sequence length and domain organization. The conserved serine-rich repeat regions SSR1 and SSR2 and the variable basic region (BR) are indicated by different colors. Gray-shaded areas connecting the sequences indicate regions of sequence identity. Sequence lengths and positions are shown in base pairs (bp).

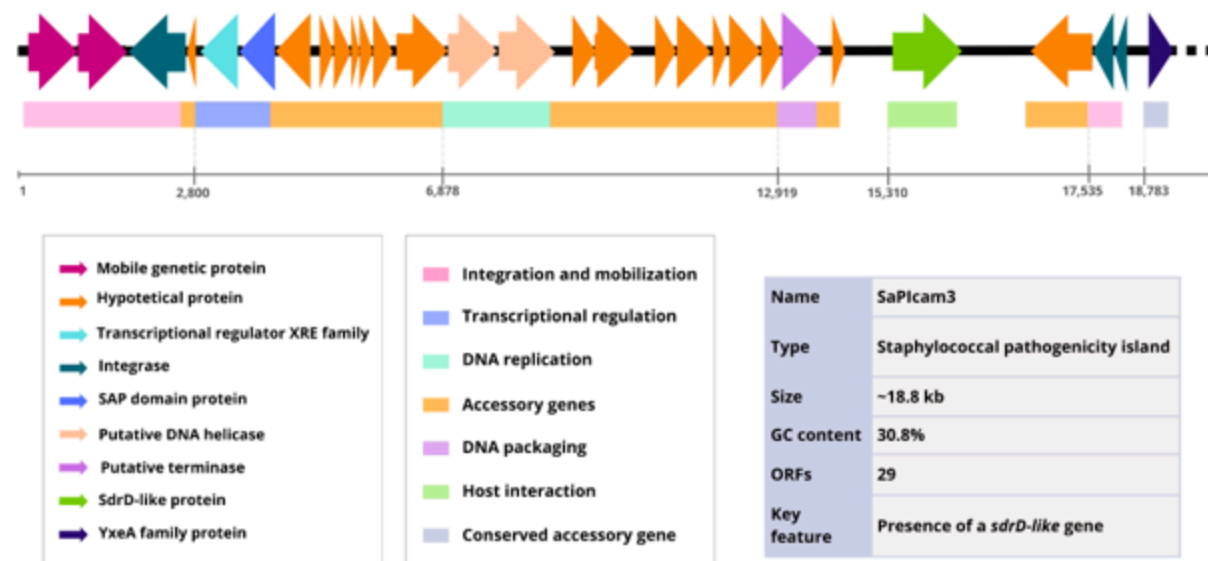


Figure 3

Page 21/23

Genomic organization of the previously undescribed SaPlcam3. Schematic representation of the SaPlcam3 region identified in the camel-associated ST30 lineage. Predicted ORFs are represented by arrows indicating their transcriptional orientation and are color-coded according to functional category, as indicated in the figure. The *sdrD-like* gene is indicated separately. Genomic coordinates correspond to the reference genome GCF_025559685.1.

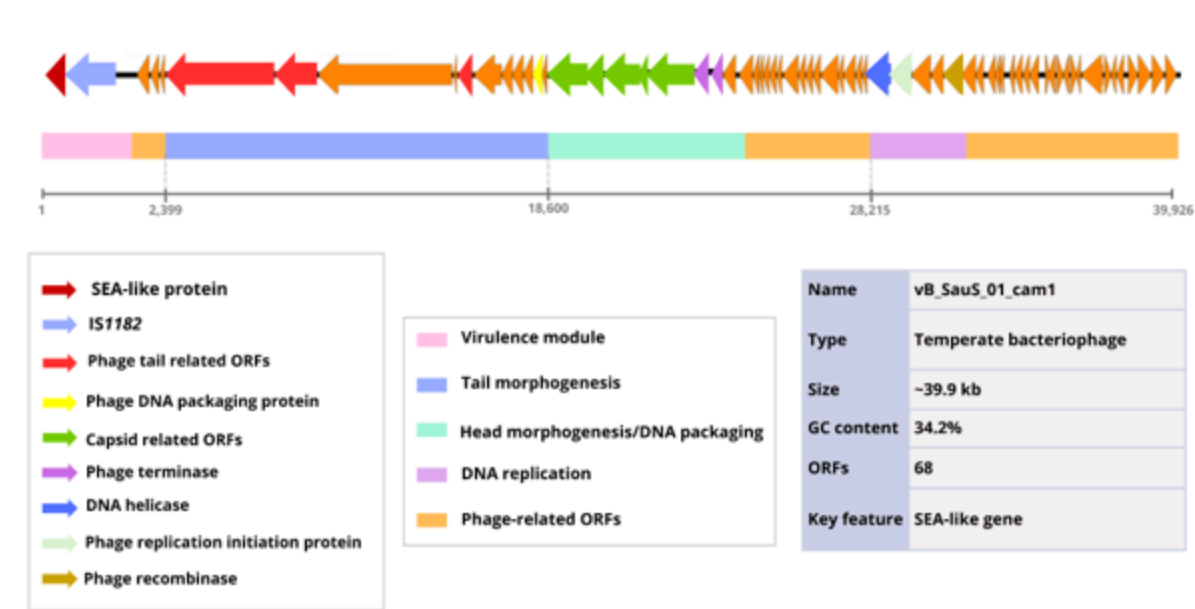


Figure 4

Genomic organization of the previously undescribed sea-like-carrying prophage vB_SauS_01_cam1. Schematic representation of vB_SauS_01_cam1, identified in this study in the camel-associated ST30 lineage. Predicted ORFs are represented by arrows indicating their transcriptional orientation and are color-coded according to functional category. The *sea-like* gene is indicated separately. Genomic coordinates correspond to the reference genome GCF_025559685.1.

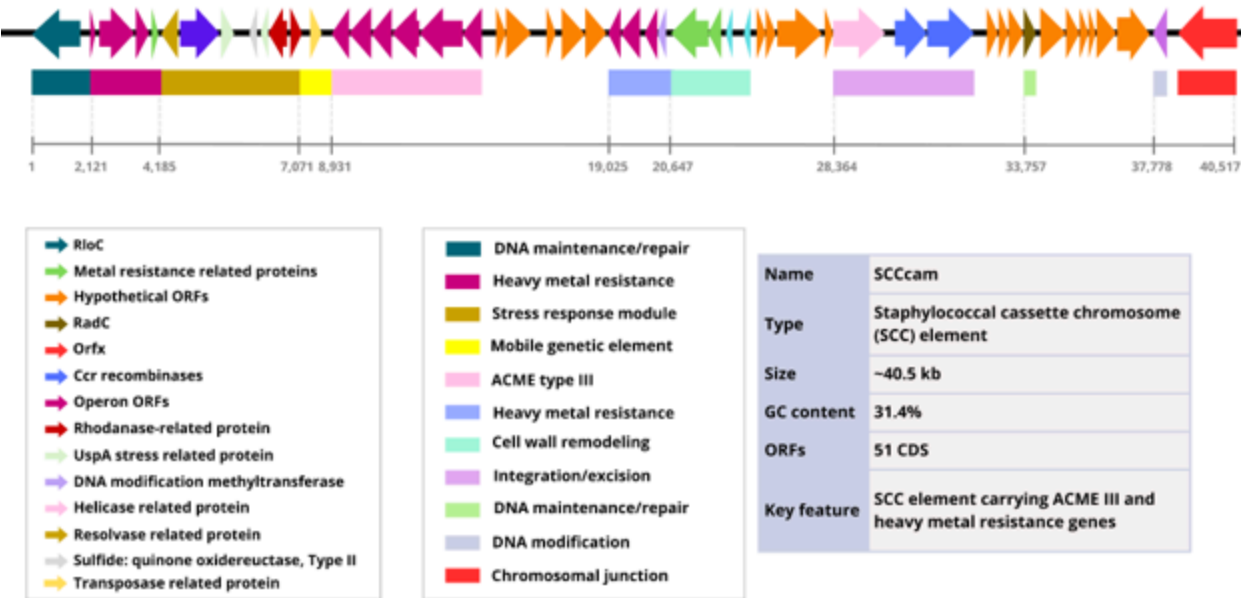


Figure 5

Genomic organization of the previously undescribed SCCcam. Schematic representation of SCCcam identified in the camel-associated ST30 lineage. Predicted ORFs are represented by arrows indicating their transcriptional orientation and are color-coded according to functional category. The ACME type III region and other functional modules are indicated in the figure. The *attL* and *attR* attachment sites delimit the SCCcam region, which is located adjacent to *rlmH* (*orfX*). Genomic coordinates correspond to the reference genome GCF_025559685.1.

Supplementary Files

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

- [Estevesetal.2026SupplementaryTable4.xlsx](#)
- [Estevesetal.2026SupplementaryTable2.xlsx](#)
- [Estevesetal.2026SupplementaryFigure1.pdf](#)
- [Estevesetal.2026SupplementaryTable6.xlsx](#)
- [Estevesetal.2026SupplementaryTable1.xlsx](#)
- [Estevesetal.2026SupplementaryTable5.xlsx](#)
- [Estevesetal.2026SupplementaryTable3.xlsx](#)
- [ExtendedDataFigs.docx](#)