

RESEARCH

Open Access



Staphylococcus aureus sequence type 71 is a chimera that emerged twice

Stefan Monecke^{1,2,3*}, Sascha D. Braun^{1,2,3}, Celia Diezel^{1,2,3}, Elke Müller^{1,2,3}, Martin Reinicke^{1,2,3}, David C. Coleman⁴, Hosny El-Adawy^{5,6}, Orla M. Keane⁷, Amira A. Moawad^{5,8}, Renata Piccinini⁹ and Ralf Ehricht^{1,2,3,10}

Abstract

Background *Staphylococcus aureus* is a Gram-positive coccus that colonises or infects humans and a wide range of domestic and wild animals. Various lineages of *S. aureus* are adapted to ruminants, including cattle, where they commonly cause mastitis. This is of significant animal welfare and economic relevance. Clonal complex (CC) 97 is one of the *S. aureus* lineages found predominantly in bovines, but also in other animals and humans. Sequence Type (ST) 71 is a related bovine-associated lineage with part of its genome identical to CC97. Thus, it has been assumed that ST71 was a chimera consisting of a fragment of DNA from an unknown donor strain inserted into a CC97 backbone genome.

Results In this study, *S. aureus* ST3042 was identified as a putative donor of this fragment based on core genome multilocus sequence typing and alignment of core and genomic island markers. We also identified a second ST71 lineage from Sardinia that differs from previously known ST71 from Ireland, the United Kingdom and continental Europe, both in size and demarcations of the ST3042-derived insert. Sardinian ST71 sequences also differed from other ST71 in the orientation of a large part of the genome. As the inversion affected the ST3042-derived region, this event must have taken place later in time than the recombination of CC97 and ST3042.

Conclusions The observation that ST71 emerged twice adds to the body of evidence that recombination involving large segments of the genome is another, highly effective, mode of horizontal gene transfer in *S. aureus* and likely contributes to the extraordinary success of this pathogen.

Keywords *Staphylococcus aureus*, Horizontal gene transfer, Chimerism, Bacterial evolution, Next-generation sequencing

*Correspondence:

Stefan Monecke
stefan.monecke@leibniz-ipht.de

¹Leibniz Institute of Photonic Technology (IPHT), Leibniz Center for Photonics in Infection Research (LPI), Jena, Germany

²InfectoGnostics Research Campus, Jena, Germany

³Center for Translational Medicine (CETRAMED), Jena University Hospital, Friedrich Schiller University Jena, Jena, Germany

⁴Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental University Hospital, University of Dublin, Trinity College Dublin, Dublin, Ireland

⁵Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institut (Federal Research Institute for Animal Health), Jena, Germany

⁶Faculty of Veterinary Medicine, Kafrelsheikh University, Kafr El Sheikh, Egypt

⁷Animal and Bioscience Department, Teagasc, Grange, Dunsany, Co. Meath, Ireland

⁸Agriculture Research Center (ARC), Animal Health Research Institute, Giza, Egypt

⁹Department of Veterinary Medicine and Animal Science - DIVAS, University of Milan, Lodi, Italy

¹⁰Institute of Physical Chemistry, Friedrich-Schiller University, Jena, Germany



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Staphylococcus aureus is a Gram-positive coccus that colonises 20–30% of any human population [1] but can also be found in domestic and wild [2, 3] animals. It is able to cause a wide range of infections, including skin and soft tissue infections (SSTIs), bone, joint and implant infections, pneumonia, septicaemia and various toxicoses. Various lineages of *S. aureus* are adapted to ruminants, including cattle, where they are a frequent etiological agent of mastitis causing a high economic burden with costs arising from veterinary treatments, animal culling, and the loss of milk production.

Mobile genetic elements play a significant role in the evolution of *S. aureus* as well as in host-adaptation. Bacteriophages are extremely important in transmitting leukocidin genes [4–7], immune evasion genes [8, 9] and the exfoliative toxin gene A [10]. Staphylococcal Chromosome Cassette (SCC) elements are large genomic islands that encode recombinase genes and variable other payloads, commonly including the beta-lactam resistance genes *mecA/C* [11–15] or the fusidic acid resistance gene *fusC* [16, 17].

Mobile elements are not the only pathway to horizontal gene transfer utilised by *S. aureus*. Large-scale genomic replacements, or chimerism, might also contribute to a horizontal flow of genetic material. Huge parts of the genome can be transferred and integrated into *S. aureus* of other, frequently distantly related, lineages by a yet unknown mechanism(s). Multilocus-sequence typing (MLST) provides a convenient tool for detecting such cases. In MLST, seven housekeeping genes are defined, and their alleles are numbered [18]. If two strains or isolates have identical profiles, they belong to the same Sequence Type (ST). Non-identical, but related STs form Clonal Complexes (CCs). If two STs share, for instance, six alleles, but significantly differ in a seventh, this might indicate a recombination event. With increasing availability of whole-genome sequencing methods (and increasing computing power), schemes have been proposed expanding the MLST approach to the entirety of the non-mobile or conserved part of the bacterial genome, i.e., on more than 2,000 genes rather than just seven [19]. This approach is known as core genome MLST (cgMLST).

Previously reported examples of chimeric lineages include ST34 and ST42 (resulting from transfer of genomic DNA from CC10 into a CC30 genome [20, 21]), a CC1 x CC80 chimera from East Africa [22], several CC133-derived chimeras from waterfowl [23], ST239 (a pandemic methicillin-resistant *S. aureus* (MRSA) clone comprising CC30 genomic DNA integrated into a CC8 backbone genome [20]), a CC9 x CC398 chimera from European livestock [24], ST617 (a CC8 x CC45 chimera [25]), ST1048 and ST1774 (chimeric MRSA from Hong Kong [26]), ST2249 (a chimeric MRSA from Australia,

derived from as many as three parental lineages: CC8, CC30 and CC45 [27]) and ST6610 (a CC8 x CC140 chimeric MRSA from East Africa [28]). ST71 is another chimeric lineage, first described from Irish livestock but also extant elsewhere in Europe [29–32]. It is clearly derived from the pandemic CC97 livestock lineage. However, it differs in several notable features, all located near the origin of replication (*oriC*). These include the absence of the histidine biosynthesis operon *his* and the biofilm-associated *ica* operon in ST71, the presence in ST71 of the collagen binding adhesin gene (*cna*) and of the enterotoxin-like gene ORF CM14 (corresponding to GenBank U10927.2, positions 32,627..33,406) or to locus tag SAB0026 in GenBank AJ938182.1; recently dubbed as *selz*: [33]). Canonical CC97 and ST71 furthermore differ in the allele of the MLST marker gene *arcC* (3 in ST97, 18 in ST71) and in alleles of other genes such as *lmrP*. They also differ in capsule type (5 in ST97, 8 in ST71). These differences are consistent with the replacement of a large fragment of more than 300,000 bp of the CC97 genome with a segment originating from a different strain [30]. In addition, a 30,000 bp fragment from this region that encompasses the *his* and *ica* loci must have been lost [30]. The origin of this inserted region could not be determined in the initial 2015 study, as its key marker, ORF CM14, is not widespread and none of the ORF CM14-positive lineages then sequenced fit the other characteristics of the segment in question [30]. A possible explanation was that the inserted region itself was a composite or chimera consisting of many small fragments of different origins [32].

The purpose of the present study was to re-assess the evolutionary history of ST71 given the increased availability of sequencing technologies and genomic data in recent years.

Materials and methods

Strains and isolates

An overview of isolates and sequences considered is provided in Table 1.

Three ST71 isolates from Ireland were subjected to long-read sequencing using Oxford Nanopore Technology (ONT). This included isolate MOK099, that was previously sequenced using Illumina short-read sequencing technology (SAMN03247575, [30]), and two new isolates (FD516C and 12_306). Five isolates from mainland Italy were also ONT sequenced. Four were previously assigned by DNA microarray profiling to ST71 (Milano-07, Milano-08, Milano-09 and Milano-84) and one could not be assigned to a ST using the array data (Milano-71). As an isolate of an “unknown *agr* IV lineage” [37], it was later subjected to conventional MLST typing [18] yielding ST3042. Two ST71 isolates (22CS0319-1 and 22CS0351) were collected from a sheep and a cow,

Table 1 Isolates/sequences, their provenience and typing data

Isolate/sequence	Reference	GenBank accession	Provenience	Isolation source	Resistance genotype	MLST	MLST profile (arcC-L2-aroE-glpF-gmk-ptx-tpi-yqiL)	RIDOM spa type	RIDOM spa profile	agr group	Cap-sule type
Milano-07	This study	Pending*	Mainland Italy, 2012	Bovine quarter milk sample	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
Milano-08	This study	Pending*	Mainland Italy, 2012	Bovine quarter milk sample	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
Milano-09	This study	Pending*	Mainland Italy, 2012	Bovine quarter milk sample	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
Milano-84	This study	Pending*	Mainland Italy, 2012	Bovine quarter milk sample	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
22CS0319-1	This study [34],	Pending*	Sardinia, 2021	Milk of mastitic sheep	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
22CS0351	This study	Pending*	Sardinia, 2022	Milk of mastitic cow	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
FD516C	This study	Pending*	Ireland, 2011	Bovine milk	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
12_306	This study	Pending*	Ireland, 2012	Bovine milk	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
MOK042	[31]	CP029627	Ireland, 2010	Bovine milk	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
MOK099	[30]	SAMN03247575, re-sequenced for this study *	Ireland, 2011	Bovine milk	mecA/C neg.	ST71	18-1-1-1-1-5-3	t528 or t524*	04 or 04-17**		8
A_SAU109	-	CP186909	Mainland Italy	Bovine milk	mecA/C neg.	ST71	18-1-1-1-1-5-3	t524	04-17		8
MOK063	[31]	CP029629	Ireland, 2010	Bovine milk	mecA/C neg.	ST97	3-1-1-1-1-5-3	t359	07-23-12-21-17-34-34-33-34		5
Wood46 (NCTC7121, NCTC10344)	[35]	LR134087, LS483324	Not recorded, earlier than 1946	Not re-corded	mecA/C neg.	ST97	3-1-1-1-1-5-3	t359	07-23-12-21-17-34-34-33-34		5
NCTC9752	-	LS483310	UK, before 1955	Not re-corded	mecA/C neg.	ST97	3-1-1-1-1-5-3	N/A	spa gene absent		5

Table 1 (continued)

Isolate/sequence	Reference	GenBank accession	Provenience	Isolation source	Resistance genotype	MLST	MLST profile (arcC-L2-aroE-glpF-gmk-pta-tpi-yqiL)	RIDOM spa type	RIDOM spa profile	agr group	Cap-sule type
FDA209P (ATCC6538, JCM2151)	[36]	CP123853, CP158284, AP043700	Not recorded	Not recorded	mecA/C neg.	ST464	3-1-1-1-29-5-3	t3297	14-12-21-17-34-34-34-33-34	I	5
NCTC4137	-	LR134091	USA, 1933	Human, compound fracture	mecA/C neg.	ST464	3-1-1-1-29-5-3	t934	07-23-12-34-34-34-34-33-34	I	5
IVRI-FBI-646	-	CP125746	India, 2019	Animal milk	mecA/C neg.	ST2219	3-1-280-168-1-5-3	t359	07-23-12-21-17-34-34-33-34	I	5
Strain-15	-	CP094925	Mexico, 2015	Milk of mastitic cow	mecA/C neg.	ST4552	3-1-1-37-1-5-611	t224	07-23-12-21-17-34-33-34	I	5
Milano-71	This study [37],	Pending*	Mainland Italy, 2012	Bovine quarter milk, taken for screening purposes	mecA/C neg.	ST3042	18-146-45-18-343-329-375	t14273	04-17-25-23-34-50-17	IV	8

*See also Supplemental file 1

**Discrepancy noted between the Illumina and Oxford Nanopore sequences, which might be a sequencing artefact, or a mutational change after many in vitro passages

respectively, on the Italian island of Sardinia. In addition, a previously published ST71 genome sequence (Irish *S. aureus* strain MOK042, GenBank accession number CP029627.1 [31]) and three genome sequences of canonical CC97 (*S. aureus* strain MOK063, CP029629.1 [31]; strain NCTC4137, LR134091.1 and strain NCTC9752, LS483310.1) were analysed in detail.

ST71 short-read sequences from the cgMLST database at https://pubmlst.org/bigdb?db=pubmlst_saureus_seqdef were downloaded to be used for comparison of particular aspects (see below). These are listed in Table 2.

For comparison, previously published CC97 genomes were included. Several sequences originated from identical strains, kept under different names by different laboratories. ATCC6538, FDA209P and JCM 2151 (GenBank accession numbers CP123853, CP158284, AP043700) are the same strain; and NCTC10344 (GenBank LS483324) is a variant of NCTC7121 (Wood46; GenBank LR134087). They were deliberately included to assess the variability of cgMLST markers as well as of the number of genes encoding tandem repeat proteins within a genomic island of relevance for this study (see below).

Nanopore sequencing

All bacterial strains were cultured overnight at 37 °C on Columbia blood agar (Becton Dickinson, Heidelberg, Germany).

Table 2 ST71 short-read sequences from the PubMLST database and their provenience

cgMLST-ID	Strain ID or Accession number	Country of origin
cgMLST-ID 50362	ASM318612v1	Ireland
cgMLST-ID 42601	Strain 682	Italy
cgMLST-ID 43035	Strain 36093	Italy
cgMLST-ID 43046	26520_14107	Italy
cgMLST-ID 42246	DS1	Austria
cgMLST-ID 42738	G071	Switzerland
cgMLST-ID 42270	DS96	Denmark
cgMLST-ID 42368	m_MR_7M	Denmark
cgMLST-ID 14446	ERR246599	UK
cgMLST-ID 14447	ERR246600	UK
cgMLST-ID 14448	ERR246603	UK
cgMLST-ID 14480	ERR246642	UK
cgMLST-ID 15071	ERR246653	UK
cgMLST-ID 10102	ERR211696	unknown
cgMLST-ID 10148	ERR211697	unknown
cgMLST-ID 10922	ERR211682	unknown
cgMLST-ID 11916	ERR175902	unknown
cgMLST-ID 11922	ERR175921	unknown
cgMLST-ID 13145	ERR175888	unknown
cgMLST-ID 13158	ERR175924	unknown
cgMLST-ID 15002	ERR234764	unknown
cgMLST-ID 25318	ERR525080	unknown

Genomic DNA for Nanopore MinION sequencing (Oxford Nanopore Technology, ONT, Oxford, UK) was extracted using the Nucleospin Microbial DNA Kit (Macherey Nagel, Düren, Germany). For each strain, one full inoculation loop was washed with 500 µL of 1× PBS (pH 7.4), centrifuged for 5 min at 4,000 × g, and resuspended in 100 µL of the kit's elution buffer (buffer BE).

The extraction procedure followed the manufacturer's protocol with two minor modifications: samples were lysed in a BeadBeater (BeadBug Homogenizer, Biozym, Germany) for 12 min at maximum speed, and proteinase K was subsequently inactivated by incubation at 70 °C for 5 min. After cooling, 4 µL of RNase (100 mg/mL; Sigma Aldrich, Steinheim, Germany) were added, and samples were incubated for 5 min at 37 °C. DNA was finally eluted in 70 µL of nuclease-free water (Carl Roth, Karlsruhe, Germany).

Following extraction, DNA was prepared for sequencing using the native barcoding kit 24 V14 (SQK-NBD114.24; Oxford Nanopore Technologies, Oxford, UK) in accordance with the manufacturer's protocol. Prior to library preparation, DNA samples underwent size selection using AMPure beads (Beckman Coulter, Krefeld, Germany) at a 1:1 (v/v) ratio. Approximately 600 ng of DNA per sample, quantified with a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), were used for library preparation before loading onto an ONT flow cell (FLO-MIN114) equipped with R10.4.1 pores.

Sequencing was performed on a MinION Mk1B device for 90 h, starting with at least 900 active pores and managed through MinKNOW software (v24.11.10). Base calling was carried out using the dorado basecaller (v1.0.2) with the model `es_dna_r10.4.1_e8.2_400bps_sup@v5.0.0_bacterial-methylation`, generating Fastq files containing 4,000 reads each. Barcode trimming was performed with dorado demux (v0.7.3) using the same model.

Quality-filtered reads were assembled into contigs using Flye software (v2.9.5; [38]). The resulting assemblies were polished in two stages: first, with four iterative rounds of Racon (v1.5.0; [39]) using the parameters match 8, mismatch 6, gap 8, and a window length of 500; followed by a final polishing step with Medaka (v1.12.1; [40]) employing the model `r1041_e82_400bps`. The final corrected assemblies were subsequently used for downstream analyses.

Results

Table 1 summarises the typing data for the nineteen isolates/sequences considered in detail. All ST71 isolates/sequences differed from canonical CC97 in their *arcC* allele (18 versus 3) and in the *spa* type, while ST3042 displayed the same *arcC* allele as ST71 and a related *spa* type. ST71 has the same capsule type as ST3042 but the

agr alleles of canonical CC97. An abridged version of the cgMLST data and of data for some genomic islands are provided in Table 3. For a complete version, see Supplemental file 2. Simplified diagrams of ST71 genomes are provided in Fig. 1. A detailed description of the cgMLST data, the recombination breakpoints and the genes affected are provided in the following paragraphs.

In short, analyses by cgMLST of the core genomes of CC97, ST71 and ST3042 indicate that ST71 is derived from CC97 by insertion of a large fragment of DNA (approx. 412,500 bp) around *oriC* that originated from a donor identical or closely related to ST3042. However, there were two clearly distinct variants of ST71 that could be distinguished based on the exact localisations of the recombination breakpoints. One variant was represented by the two Sardinian isolates, while all other published and study sequences represent the other variant. One recombination breakpoint was situated within the *pfoR* (SAUR0307) gene in the majority of ST71 sequences. However, in the Sardinian sequences it bisected a gene (*esat6-11=esaD=essD*) that belongs to a genomic island around SAUR0300. The other recombination breakpoint was, in the majority of sequences, situated in the *fda* (SAUR2817) gene but could be found within *acrR=gbaA* (SAUR2789) of the Sardinian isolates. This suggested that the Sardinian ST71 emerged independently from the other ST71, by another recombination event involving the same parental strains.

All ST71 isolates lacked the genes from *ywrF* (SAUR2882) to *pcp* (SAUR2913). This deletion included the *ica* (SAUR2888 to -2892) and *his* (SAUR2895 to -2903) operons as noted previously [30]. The ST3042 sequence showed a similar loss, although even more genes in this region, including the collagen adhesin gene *cna* (which is present in ST71 but absent in canonical CC97), were replaced by transposase genes (see below).

In the two Sardinian sequences, an inversion of a large part of the genome was observed. This affected a region between SAUR0106/SAUR0107 and SAUR2880/SAUR2914, i.e., of as much as 2.6 Mbp.

Core genomic markers

Core genomic markers in *S. aureus* (and in related species *S. argenteus*, *S. schweitzeri* and *S. roterodami*) usually follow a fixed order, in which the numbers of the cgMLST markers from SAUR0001 to SAUR2939 essentially denote their relative position in the genome. cgMLST markers from SAUR2940 upwards were apparently included later in the scheme and, hence, they are scattered across the genome.

The numerical system of cgMLST markers provides a concise overview on the identity of genomic regions and facilitates a quick comparison of canonical CC97, ST71 and ST3042 sequences (see Table 3 for an abridged

version and Supplemental file 2 for the full version). The only drawback is the fact that there are no other published genome sequences for ST3042 than the one described herein. Thus, some alleles may not be represented in the MLST database; however, if novel alleles are identified, the genes will at least be marked as “present” in Table 3, and/or alignments are provided as Supplemental files 3 and 4 to facilitate a direct comparison. There are also multiple alleles known to be associated with CC97, making it easier to find or to rule out a match to ST71 sequences. For instance, *wall=yyeI* (SAUR0023) can have alleles 209 or 211 in CC97, 59 in ST71, but a novel allele is present in the single ST3042 genome sequenced, with no number yet assigned. In all ST71 sequences apart from two, cgMLST alleles from *dnaA* (SAUR0001) to A5IPH6 (SAUR0306) matched alleles from ST3042 while from Q5HJ57 (SAUR0308) until close to the end of the genome they correspond to CC97 (with *pfoR* (SAUR0307) having a unique allele, see below). The two ST71 exceptions were the two Sardinian isolates in which *dnaA* (SAUR0001) to *esat6-05=essB* (SAUR0285) match ST3042 while they corresponded to CC97 from DUF0600-TIGR01741_MW0279 (SAUR0298) onwards. This placed the putative recombination breakpoint for these two isolates into a genomic island as discussed below.

For the markers downstream of SAUR0298 (Sardinian ST71 isolates) or SAUR0308 (other ST71 isolates), alleles in ST71 correspond to those in canonical CC97.

However, the “last” CC97-like gene in the two Sardinian isolates was Q5HCX4=DUF2316 (SAUR2788). The gene *acrR=gbaA* (SAUR2789) differed from both CC97 and ST3042 sequences (see below). From SAUR2790 onwards, alleles were identical or related to ST3042 as the alignment shows (see below) although there are no cgMLST allele numbers yet assigned. In all other ST71 sequences, Q5HCU7 (SAUR2816) is the last gene with an allele matching CC97. The *fda* (SAUR2817, fructose-bisphosphate aldolase class 1) allele was novel, present neither in CC97 nor in ST3042. From *lqo* (SAUR2818) downstream to *oriC* alleles match ST3042.

The absence of two major gene loci from the regions between SAUR2788/2816 and *oriC* of ST71 and ST3042 will be discussed below.

Analysis of genomic islands

S. aureus harbours a large number (approximately 80–100) of genomic islands (GIs), i.e., genes or clusters of genes that are frequently, but not universally present, that always inhabit the same position within the genome and whose presence is strictly linked to the clonal complex affiliation. Some GIs consist of only one gene or a small number of genes, while more than 50 different genes might be associated with some of the major GIs (although not all of them are simultaneously

Table 3 cgMLST data and some genomic islands (abridged version, for full version see Supplemental file 2). The analysis is based on the strains/sequences listed in Table 1 but the sequences from the PubMLST databases were included into underlying alignments (see Supplemental files 3 and 4). Genes are listed in the standard order, but the inverted region in Sardinian isolates is left-aligned and marked with green font. Numerical codes refer to cgMLST profiles (https://pubmlst.org/bigsdb?db=pubmlst_saureus_seqdef), except for the *spa* types (RIDOM) and the seven classical MLST markers according to Enright [18]; these are marked “MLST-7” in bold font. “Present” means cgMLST markers where the gene was detected but no code number was yet assigned. For genomic island (GI) markers, it just indicates presence. N/A means complete absence or (theoretically) truncation beyond recognition. Suppl. 3/4, S3 and S4; Sequences/alignments are provided in Supplemental files 3 and 4, respectively. Blue shade indicates CC97 provenience; yellow, ST3042; and grey, undetermined provenience (based on cgMLST or alignments in Supplemental files 3 and 4). Genes shaded red are putative chimeric genes, see the alignments in Figs. 2, 3, 4 and 5 and Supplemental files 3 and 4 [41]

Gene ID	Class	Suppl. 3/4	cgMLST	CC97 (n=7 strains/10 sequences)	ST71 (n=9)	Sardinian ST71 (n=2)	ST3042 (n=1)
<i>dnaA</i>	Core	-	SAUR0001	25 or 298	77	Present	Present
<i>dnaN</i>	Core	-	SAUR0002	31	93	Present	Present
<i>yaaA</i>	Core	-	SAUR0003	1	3	3	3
<i>recF</i>	Core	-	SAUR0004	7,21,54,263 or 474	76	346	Present
<i>gyrB</i>	Core	-	SAUR0005	38 or 710	118 or 318	118	118
...							
<i>walI=yycI</i>	Core	-	SAUR0023	209 or 211	55	55	Present
<i>walJ=yycJ</i>	Core	-	SAUR0024	22	27	27	27
<i>sasH=adsA</i>	Core	-	SAUR0025	239 or 286	65 or 189 or 1263	Present	Present
<i>orfX</i>	Core	-	SAUR0026	52 or 103	144	144	144
GOLPR5/GOLPR6	GI	-	N/A	Present *	N/A	N/A	N/A
<i>hsdM/S/R-GI</i>	GI	-	N/A	Present *	N/A	N/A	N/A
<i>entCM14</i>	GI	-	N/A	N/A	Present	Present	Present
<i>cstR/A/B-GI</i>	GI	-	N/A	Present	Present	Present	Present
<i>Q2YUT2=yut2</i>	Core	-	SAUR2960	10 *	57	57	57
...							
DUF1648	Core	-	SAUR0106	21 *	40	Present	40
<i>lctP-locus1</i>	Core	-	SAUR0107	34 *	100	100	100
<i>spa</i>	Core	-	RIDOM	07-23-12-21-17-34-34-33-34 07-23-12-34-34-34-33-34 07-23-12-21-17-34-33-34 14-12-21-17-34-34-34-33-34 *	04 04-71	04-71	04-71-25-23-34-15-17
...							
<i>deoB</i>	Core	-	SAUR0138	22 or 427	72	72	72
<i>phnE2</i>	Core	-	SAUR0139	16, 26 or 505	29	29	29
<i>phnE1</i>	Core	-	SAUR0140	35 or 466	57	57	57
<i>phnC</i>	Core	-	SAUR0141	27 or 59	60	Present	60
<i>phnB</i>	Core	-	SAUR0142	11 or 556	55	55	55
...							
<i>capG</i>	Core	-	SAUR0154	40 or 521	43 or 247	43	43
<i>capH</i>	Core	-	N/A	<i>capH</i> -5	<i>capH</i> -8	<i>capH</i> -8	<i>capH</i> -8
<i>capJ</i>	Core	-	N/A	<i>capJ</i> -5	<i>capJ</i> -8	<i>capJ</i> -8	<i>capJ</i> -8
<i>capK</i>	Core	-	N/A	<i>capK</i> -5	<i>capK</i> -8	<i>capK</i> -8	<i>capK</i> -8
<i>capL</i>	Core	-	SAUR0159	34 or N/A	97	Present	Present
<i>capM</i>	Core	-	SAUR0160	19	39	39	39
...							
<i>coa</i>	Core	-	Type according to [41]	VII c	VII b	VII b	VII b
...							
Q5HJ97	Core	S3	SAUR0277	30 or 283	76 or 387	Present	Present

Table 3 (continued)

Gene ID	Class	Suppl. 3/4	cgMLST	CC97 (n=7 strains/10 sequences)	ST71 (n=9)	Sardinian ST71 (n=2)	ST3042 (n=1)
QSHJ96	Core	S3	SAUR0278	30 or 54	101	Present	Present
QSHJ95	Core	S3	SAUR0279	37 or 840	126 or 844	Present	Present
Q2G192-1 (SAA6159_00259)	GI	-	N/A	Present	N/A	N/A	N/A
Q2G192-2 (SAMSHR1132_02470)	GI	-	N/A	Present	N/A	N/A	N/A
Q6GK17 (SACOL0269)	GI	-	N/A	Present	N/A	N/A	N/A
<i>ssaA=saCf</i>	GI	S3	SAUR0280	Present	Present	Present	Present
<i>esat6-01=esxA</i>	GI	S3	SAUR0281	4	1	1	1
<i>esat6-02=esaA</i>	GI	S3	SAUR0282	43	201 or 1548	Present	Present
<i>esat6-03=essA</i>	GI	S3	SAUR2978	4	4 or 80 or 179	4	4
<i>esat6-04=esaB</i>	GI	S3	SAUR0284	70	2	2	2
<i>esat6-05=essB</i>	GI	S3	SAUR0285	370 or 664	97	97	Present
<i>esat6-06=essC</i>	GI	S3	SAUR0286	Present	Present	Present	Present
<i>esat6-07=esxC=esaC</i>	GI	S3	N/A	Present	Present	Present	Present
<i>esat6-08=esxB</i>	GI	S3	N/A	Present	Present	Present	Present
<i>esat6-09=esaE=DUF5081</i>	GI	S3	N/A	Present	Present	Present	Present
<i>esat6-10=esxD</i>	GI	S3	N/A	Present	Present	Present	Present
<i>esat6-11=esaD=essD</i>	GI	S3	N/A	Present	Present	Chimeric	Present
DUF0600-TIGR01741_esa6	GI	S3	N/A	Present (SACOL0282-like)	Present (MW0269-like)	Present (SACOL0282-like)	Present (MW0269-like)
Q1Y4R2=DUF5079	GI	S3	N/A	Present *	Present	Present	Present
Q1Y4R3=DUF5080	GI	S3	N/A	Present *	Present	Present	Present
Q9L3N6	GI	S3	N/A	Present *	N/A	Present	N/A
DUF0600-TIGR01741	GI	S3	N/A	3 copies *	1 copy as in ST3042	3 copies as in most CC97	Present
Q1Y4R2=DUF5079	GI	S3	N/A	Present	N/A	Present	N/A
Q1Y4R3=DUF5080 (truncated)	GI	S3	N/A	Present	N/A	Present	N/A
DUF0600-TIGR01741	GI	S3	N/A	1 to 3 copies	N/A	1 copy	N/A
AD196784=DUF5083	GI	S3	N/A	Present or N/A **	Present or N/A	N/A	Present
DUF4467-CA347_306	GI	S3	N/A	Present or N/A **	Present or N/A	N/A	Present
DUF4064-GI	GI	S3	N/A	Present or N/A **	Present	N/A	Present
DUF5084	GI	S3	N/A	Present or N/A **	Present	N/A	Present
DUF0600-TIGR01741	GI	S3	N/A	1 to 3 copies	3 copies as in ST3042	1 copy as in MOK063, CP029629.1	3 copies
TIGR01741_MW0279	GI	S3	SAUR0298	16	Present	16	Present
Q1Y4S0=DUF4467	Core	S3	SAUR2979	14	31	14	31
Q2G173=DUF4064	Core	S3	SAUR0300	12 or N/A	42	12	42
<i>focA-L1=nirC</i>	Core	S3	SAUR0301	66	95	66	95
<i>brnQ2</i>	Core	S3	SAUR0302	14 or 16 (1308 bp, no frameshift)	545 or 548 or 705 (1307 bp, with frameshift)	14 (1308 bp, no frameshift)	Present (1307 bp, with frameshift)
USA300HOU_0326	Core	S3	N/A	Present	Present	Present	Present
Q99WS0=sap5	Core	S3	SAUR0303	29	78 or 516	29	78
SAUR0304	Core	S3	SAUR0304	1	29	1	29
Q2G168	Core	S3	SAUR0305	24	71	24	71
ASIPH6	Core	S3	SAUR0306	24 or 319	28	24	28
USA300HOU_0326	Core	S3	N/A	Present	Present	Present	Present
<i>pfoR</i>	Core	S3	SAUR0307	20	74 (chimeric)	20	Present
QSHJ57	Core	S3	SAUR0308	30	30	30	Present
<i>psuG</i>	Core	S3	SAUR0309	22	22 or 132	22	Present
<i>nupC2</i>	Core	S3	SAUR0310	161 or 188	161	161	Present
<i>nanT</i>	Core	S3	SAUR0311	311	311	311	Present
<i>nanA</i>	Core	S3	SAUR0312	23	23	23	Present
<i>nanK</i>	Core	S3	SAUR0313	24	24	24	Present

Table 3 (continued)

Gene ID	Class	Suppl. 3/4	cgMLST	CC97 (n=7 strains/10 sequences)	ST71 (n=9)	Sardinian ST71 (n=2)	ST3042 (n=1)
<i>nanR=rpIR</i>	Core	S3	SAUR0314	23, 307 or 418	23	23	Present
<i>nanE</i>	Core	S3	SAUR0315	12 or 55	12	12	Present

<i>yqil</i>	Core	-	(MLST-7)	3 or 611	3	3	375

<i>pta</i>	Core	-	(MLST-7)	1 or 29	1 or 459	1	343

<i>tpi</i>	Core	-	(MLST-7)	5	5	5	329

<i>gmK</i>	Core	-	(MLST-7)	1, 37 or 168	1	1	18

<i>glpF</i>	Core	-	(MLST-7)	1 or 280	1	1	45

<i>aroE</i>	Core	-	(MLST-7)	1	1	1	146

<i>hld</i>	Core	-	SAUR3002	1	1	1	Present
<i>agrB</i>	Core	-	SAUR2953	2	2	2	296
<i>agrD-I</i>	Core	-	SAUR2954	1	1	1	N/A
<i>agrD-IV</i>	Core	-	SAUR2957	N/A	N/A	N/A	1
<i>agrC</i>	Core	-	SAUR2958	6, 40, 721 or 903	194	40	954
<i>agrA</i>	Core	-	SAUR2959	53 or 624	53	53	Present

QSHCX6	Core	-	SAUR2786	6	6	6	2
<i>nmrA</i>	Core	S4	SAUR2787	7	7	7	2
QSHCX4=DUF2316	Core	S4	SAUR2788	30 or 201	30	30	58
<i>acrR=gbaA</i>	Core	S4	SAUR2789	35	35	Chimeric	11
<i>gbaB</i>	Core	S4	SAUR2790	25	25	Present	207
QSHCX0	Core	S4	SAUR2791	33 or 41	33	Present	Present
<i>yraK</i>	Core	S4	SAUR2792	11 or 527	11	Present	N/A
<i>cobW3</i>	Core	S4	SAUR2793	11	11	Present	Present
Q6GDL8=feoB	Core	S4	SAUR2794	25 or 508	25	Present	Present
QSHCW6	Core	S4	SAUR2795	16	16	Present	Present
DUF4176=esaC	Core	S4	SAUR2796	12	12	2	2

<i>alsD-L2=budA</i>	Core	S4	SAUR2811	24	24	9	9
<i>ldh2</i>	Core	S4	SAUR2812	25	25	61	Present
SAUR2813	Core	S4	SAUR2813	2	2	2	2
<i>yfnA1=pheP</i>	Core	S4	SAUR2814	35 or 634	35	Present	Present
<i>yhxA=gabT</i>	Core	S4	SAUR2815	16 or 743	16	Present	Present
QSHCU7	Core	S4	SAUR2816	7 or 296	7	Present	Present
<i>fda</i>	Core	S4	SAUR2817	117	33 (chimeric)	53	Present
<i>lqo</i>	Core	S4	SAUR2818	29	81	81	81
USA300HOU_2601	Core	S4	SAUR2819	4	2	2	2
<i>acsA2=bclA</i>	Core	S4	SAUR2820	44 or 568	115	Present	Present
QSHCU3	Core	S4	SAUR2821	19	67	Present	67

<i>arcC-L2</i>	Core	-	(MLST-7)	3	18	18	18

QSHCN9	Core	-	SAUR2881	11	39	39 or N/A	39
<i>icaR</i>	Core	-	SAUR2888	12	N/A	N/A	N/A
<i>icaA</i>	Core	-	SAUR2889	33	N/A	N/A	N/A
<i>icaD</i>	Core	-	SAUR2890	2	N/A	N/A	N/A
<i>icaB</i>	Core	-	SAUR2891	23	N/A	N/A	N/A

Table 3 (continued)

Gene ID	Class	Suppl. 3/4	cgMLST	CC97 (n=7 strains/10 sequences)	ST71 (n=9)	Sardinian ST71 (n=2)	ST3042 (n=1)
<i>icaC</i>	Core	-	SAUR2892	12 or 558	N/A	N/A	N/A

<i>hisI</i>	Core	-	SAUR2895	20 or 254	N/A	N/A	N/A
<i>hisF</i>	Core	-	SAUR2896	52 or 380	N/A	N/A	N/A
<i>hisA</i>	Core	-	SAUR2897	12	N/A	N/A	N/A
<i>hisH</i>	Core	-	SAUR2898	17	N/A	N/A	N/A
<i>hisB</i>	Core	-	SAUR2899	42	N/A	N/A	N/A
<i>hisC1</i>	Core	-	SAUR2900	124 or 317	N/A	N/A	N/A
<i>hisD</i>	Core	-	SAUR2901	102, 103, 438 or 743	N/A	N/A	N/A
<i>hisG</i>	Core	-	SAUR2902	76, 224 or 355	N/A	N/A	N/A
<i>hisZ</i>	Core	-	SAUR2903	25	N/A	N/A	N/A

<i>pcp</i>	Core	-	SAUR2913	63	N/A	N/A	N/A
<i>padR</i>	GI	-	N/A	N/A	Present	Present	N/A
<i>Q2FUS3=bstA</i>	Core	-	SAUR2914	21	50	Present	N/A
<i>cna</i>	GI	-	N/A	N/A	Present	Present	N/A
<i>Q7WS08=citT</i>	Core	-	SAUR2916	34 or 719	112	112	N/A
<i>rarD</i>	Core	-	SAUR2917	14	83	83	N/A
<i>Q5HCK1</i>	Core	-	SAUR2918	Present	53	53	N/A
<i>nixA</i>	Core	-	SAUR2919	50	85	85	N/A

<i>cspC-L3=cspB</i>	Core	-	SAUR2929	1 or 8	1	1	1
<i>Q5HC19=immR</i>	Core	-	SAUR2930	9	52	52	52

<i>mnxE=trmE</i>	Core	-	SAUR2937	13 or 101	89	89	89
<i>rnpA</i>	Core	-	SAUR2938	6	39	39	39
<i>rpmH</i>	Core	-	SAUR2939	2	2	2	2

Absent from CC97 strain MOK063, GenBank CP029629.1

* Absent from CC97 strain NCTC9752, GenBank LS483310.1

present). The actual composition and gene content of these major GIs depend on the clonal complex affiliation. For instance, CC5 carries the *egc* cluster and *lukD/E*, CC30 has *egc* but no *lukD/E*, CC1 and CC8 are positive for *lukD/E* and negative for *egc* while CC398 lacks both, *egc* and *lukD/E*. Thus, the presence and gene content of major GIs can be used to determine the provenience of fragments of a staphylococcal genome when comparing a fragment from a putative chimeric strain(s) to the corresponding regions of the genome of putative donors.

One GI, which is of relevance for the CC97/ST71/ST3042 issue, is located directly downstream of *orfX* (SAUR0026) and the SCC integration site (Fig. 1). It is commonly associated with the exotoxin genes *seh* and ORF CM14. Excluding genes that occur in varying positions, or in multiple copies, there are about 50 genes associated with this GI (see Supplemental file 2), whose presence or absence can be used to extract information on the provenience of the genomic region. The presence of ORF CM14 was one of the first recognised features distinguishing ST71 from canonical CC97. Indeed, ST3042 also carries this gene, as well as the 15 others that comprise this GI in ST71 (although two of them,

DUF3125 and a staphylococcal interspersed repeat unit, are also present in CC97, although in other positions). Canonical CC97 lacks ORF CM14 but has another five genes (including a type I restriction-modification system) that are absent from ST71 and ST3042. Another nine genes (including the *cstR/A/B* operon) are present in CC97, ST71 and ST3042 sequences. Of note, one CC97 sequence (NCTC9752, LS483310.1; Supplemental file 2) lacks this GI and the next GI downstream as well as the *spa* gene.

A second GI, directly downstream of *dusC* and accompanying genes (see Supplemental file 2), also has the same gene content in ST71 and ST3042, which differs from the corresponding GI in CC97. Some minor islands will not be discussed here in detail but are shown in Supplemental file 2. Another major GI further downstream comprises a cluster of eleven genes encoding an ESA6-like secretion system (Ess), and some of the genes were recently included into the cgMLST scheme (SAUR0280, SAUR0281, SAUR0282, SAUR2978, SAUR0284, SAUR0285, SAUR0286). Another approximately 20 genes might, or might not, be present in this GI, depending on CC affiliation. In addition, this GI includes multiple



Fig. 1 The structure of a typical ST71 (left) and a Sardinian ST71 (right) genome. Regions of CC97 provenience are shown in blue and those of ST3042 origin are shown in yellow. Inverted regions are shown in a lighter shade and are indicated by the arrows in the inner circle. Typing markers (see Table 1) are shown in darker shade. Recombination breakpoints are indicated in red and with Roman numerals; inversion breakpoints in green. The normal position of the (absent) *his* and *ica* loci is shown in grey. The position of the SaPI with *bap* (present only in two Irish isolates) is indicated in purple

copies of a putative gene DUF0600-TIGR01741, related to MW0279, whose numbers, actual length and sequence might be variable. The Sardinian ST71 sequences matched, in this regard, one of the CC97-associated variants (MOK063) while the other ST71 sequences corresponded to ST3042 (Table 3/Supplemental file 2). An alignment of the genes (Figs. 2 and 3; Supplemental file 3) shows that one breakpoint of the large chromosomal recombination bisects one of the genes (*esat6-11=esaD=essD*) belonging to this GI, but only in the two Sardinian genomes. The alignment confirms that the exact localisation of this breakpoint in the Sardinian sequences differs from the corresponding breakpoint localisation in all other ST71 sequences (including genome sequences from the cgMLST databases). More details are discussed below.

All following GIs of ST71 either match the corresponding GIs in CC97 or are identical in all three complexes concerned. Thus, they shall not be discussed in detail here, their gene content is summarised in Supplemental file 2. Differences emerge only downstream of the second recombination breakpoint and due to large-scale deletions in ST71 and ST3042 (a detailed discussion of both issues is provided below).

The inversion of a part of the genome in Sardinian strains

The only exceptions from the strict order of core genomic markers mentioned above are occasional inversions, in which a part of the genome was excised and

re-integrated in reverse order and orientation, and hence the cgMLST code numbers are “counting backwards”. This has been described for a Russian MRSA strain (GenBank: AP017377) in which the inverted region is flanked by IS256 elements [42] and for a PVL-positive *S. argenteus*, in which the inverted region lies between two lysogenic prophage genomes [43]. Short-read (e.g., Illumina) sequencing cannot be expected always to be able to identify this phenomenon, because the natural orientation of the contigs is lost and because the flanking prophage or transposon sequences might be present in multiple copies causing contigs to end exactly at those positions that are of interest.

CC97-MSSA strain MOK063 (GenBank CP029629.1) also shows such a phenomenon, in which *lysS* (SAUR0525) is followed by SAUR2346, SAUR2345, SAUR2344 etc. until SAUR0541 after which *salA=parA* (SAUR2357) and the rest of the genome can be found in normal orientation. Here, neither prophages nor IS256 elements are noted to flank the inversion, which comprises approximately 1.67 Mbp.

The two Sardinian ST71 isolates carried such an inversion (Fig. 1). It started from SAUR0106 as the last regularly oriented and SAUR2880 (in 22CS0319-1)/SAUR2881 (in 22CS0351) as the first inverted cgMLST marker. It ended with *lctP-1* (SAUR0107) as the last inverted gene and *padR* (situated on a GI)/*bstA=Q2FUS3* (SAUR2914) as the first regularly orientated markers downstream, encompassing as much as



Fig. 2 Alignment of *esat6-11=esaD=essD* from the CC97 sequences, from Sardinia ST71 (marked as "Sard."), other ST71 and from ST3042. Alignments for this gene (and the gene shown in the following figure) that also include surrounding genes and the ST71 sequences from the cgMLST database are presented in Supplemental file 3



Fig. 3 Alignment of *pfoR* (SAUR0307) from the CC97 sequences, from Sardinian ST71 (marked as "Sard."), other ST71 and from ST3042

2.6 Mbp. Towards DUF1648 (SAUR0106), the inversion was in both strains flanked by four copies of an IS431 transposase plus two genes that usually are plasmid-associated. In isolate 22CS0319-1, a mobile genetic element of 14,000 bp harbouring 17 genes was additionally integrated there that removed Q5HCN9 (SAUR2881). Genes of this element were also present in other strains (ST71; Milano-84, MOK099 and CP186909 and CC97; LR134091), although in other positions. Towards *padR* and SAUR2914, the inversion was flanked by one copy of an IS431 transposase (in isolate 22CS0351) or, respectively, by three copies of an IS431 transposase plus the (normally plasmid-borne) *aurR/I/D/C/B/A/T* aureocin locus (22CS0319-1). Thus, a Sardinian ST71 genome basically can be regarded as an assembly of four distinct fragments. First, there was a ST3042-derived region with normal gene order and orientation, from *padR* and

SAUR2914 through *oriC*, to SAUR0106. It was followed by another ST3042-derived fragment, in inverted order and orientation, from SAUR2881 to SAUR2790, with the "second" recombination breakpoint (see below) bisecting SAUR2789. The third fragment then consisted of genes of CC97 provenience, that again are inverted, "counting backwards" from SAUR2788 to SAUR0300 and into the *ess* GI delineated by the "first" recombination breakpoint. This was followed by another, inverted region of ST3042 descent, from GI associated *esat6-10=esxD* and SAUR0286 to *lctP-1* (SAUR0107), after which, with *padR* and *bstA*, SAUR2914 the regularly orientated ST3042 region started.

Recombination breakpoint I

As mentioned above, one recombination breakpoint or fault line affects the major GI further comprising

the *ess* gene cluster. Aligning these genes, it becomes apparent that among genes that occur in both CC97 and ST3042 genomes, specific alleles can be distinguished. This was easily visible for *ssaA*=*scaF*, *esat6-01*=*esxA*, *esat6-02*=*esxA*, *esat6-03*=*essA*, *esat6-04*=*esAB*, *esat6-05*=*essB*, *esat6-06*=*essC* as these genes are part of the cgMLST scheme (SAUR0280 to SAUR0282, SAUR2978, SAUR0284 to SAUR0286) yielding, when already assigned, different allele numbers (Table 3/ Supplemental file 2), with ST71 isolates sharing alleles with ST3042, but differing from canonical CC97. This was also true for *esat6-07*=*esxC*=*esAC*, *esat6-08*=*esxB*, *esat6-09*=*esAF*=DUF5081 and *esat6-10*=*esxD*. For *esat6-11*=*esAD*=*essD*, another phenomenon could be observed (for an alignment, see Fig. 2/Supplemental file 3). Here, all ST71 sequences shared 44 single nucleotide polymorphisms (SNPs) with the ST3042 sequence, differing from canonical CC97, up to position 930 in this gene. Henceforth, Sardinian sequences were identical to CC97. All other ST71 sequences (including those from the cgMLST database) carried at positions 946/947 a double “T” which is neither present in CC97 nor in ST3042 and from a SNP in position 1024 on, they shared another 21 SNPs with ST3042, differing from the CC97 allele. For all genes downstream, the two Sardinian isolates had the same alleles as canonical CC97, while the other ST71 sequences still matched ST3042. This was also apparent by checking the cgMLST alleles for SAUR2979 and SAUR0300 to SAUR0306 (Table 3/Supplemental file 2).

CC97 and Sardinian ST71 had cgMLST allele 20 of the *pfoR* gene (SAUR0307), all other ST71 had allele 74, while for ST3042, there was no number yet assigned.

SAUR0307-20 differed from SAUR0307-74 and the ST3042 allele in 4 SNPs upstream of position 850, while SAUR0307-20 and -74 were identical henceforth, differing from ST3042 in 6 SNPs (see Fig. 3/Supplemental file 3). All following genes downstream in ST71 matched canonical CC97.

Recombination breakpoint II

Again, the exact location of the second breakpoint in the two Sardinian isolates is different than in all other ST71 sequences, even when ignoring the issue of orientation.

The sequences of the Sardinian isolates presented with an unassigned allele for the cgMLST marker *acrR=gbaA* (SAUR2789), encoding glucose-induced biofilm accessory regulator A. All other ST71 and CC97 carried SAUR2789-35 while ST3042 had SAUR2789-11. An alignment (see Fig. 4/Supplemental file 4) showed 30 SNPs separating all CC97 and ST71 sequences from ST3042; however, the Sardinian strains and ST3042 shared a SNP in the 3' end at position 480. Differences between non-Sardinian ST71 and CC97 accumulated only downstream of *fda* (SAUR2817). CC97 had SAUR2817-117. Sardinian ST71 had SAUR2817-53 and for ST3042, no allele number was yet assigned. Non-Sardinian ST71 had allele SAUR2817-33. Upon inspection of the alignment (see Fig. 5/Supplemental file 4), this allele can be regarded as chimeric with ST71 and CC97 sequences being identical for the 5' terminal positions. From position 372 to the end (at pos. 891), 30 SNPs separated ST71 from CC97, but only one from ST3042. All following genes matched, if present, ST3042 rather than CC97.



Fig. 4 Alignment of *acrR=gbA* (SAUR2789) from the CC97 sequences, from Sardinian ST71 (marked as "Sard."), other ST71 and from ST3042. Alignments for this gene (and the gene shown in the following figure) that also include surrounding genes and the ST71 sequences from the cgMLST database are presented in Supplemental file 4



Fig. 5 Alignment of *fda* (SAUR2817) from the CC97 sequences, from Sardinian ST71 (marked as "Sard."), other ST71 and from ST3042

Gene deletions in ST71 and ST3042 (*cna*, *ica* and *his* operons)

The absence of the *ica* (SAUR2888 to -2892) and *his* (SAUR2895 to -2903) operons from ST71 has been noted previously, and its histidine auxotrophy has been demonstrated [30]. Sequencing of additional isolates for the present study, including Sardinian isolates, revealed that loss of both operons is a common feature of all ST71 isolates sequenced. Indeed, they lack all genes from *ywrF* (SAUR2882) to *pcp* (SAUR2913), with the first following genes being the genomic island gene *padR* as well as *bstA*=Q2FUS3 (SAUR2914). In the Sardinian sequences, this also marked one flank of the genomic inversion described above. This region was replaced by an IS431 insertion element with a variable number of transposase genes that encoded a toxin from a plasmidic toxin/antitoxin system and one or two mobile element associated genes (corresponding to SAB1296c, SAB1297 in AJ938182.1) plus, in one case (Milano-84), an integrated plasmid. Similarly to ST71, the ST3042 isolate lacked *ica* and *his* operons. Furthermore, it also lacked the genes situated between these loci and Q6GDA0 and

cspC-L3=*cspB* (SAUR2929). This included the collagen adhesin gene *cna* (which is present in ST71 but absent in canonical CC97). Here, again, three IS431 transposase gene copies, the toxin gene and SAB1296c were present. Whether this is representative for ST3042, or for a sub-lineage within ST3042, or just an accidental deletion affecting this particular isolate cannot be determined due to a lack of sequences for ST3042 in publicly available sequence databases.

Other genes of interest

As mentioned above, one Sardinian isolate (22CS0319-1) carried the *aurR/I/D/C/B/A/T* aureocin locus [44] located on an integrated plasmid.

Two Irish isolates (FD516C_150525 and 12_306) were positive for the *bap*-gene encoding a biofilm-associated protein [45, 46] on a *S. aureus* Pathogenicity island (SaPI) located next to *guaA* (SAUR0412). The ST3042 isolate, some of the other ST71 isolates and one CC97 reference sequence (IVRI-FBI-646, GenBank CP125746) harboured another SaPI at the same position (see Supplemental File 2). This SaPI comprised, among genes for "putative

proteins” and an integrase gene, also genes encoding adenosine deaminase (*add*), a veterinary paralogue of complement inhibitor SCIN family protein (*scn2*) and a van Willebrand factor binding protein (*vwb3*; corresponding to SAPIG0481 to SAPIG0483 in GenBank AM990992.1).

Discussion

Analysis of the genomes of CC97, ST71 and ST3042 provides convincing evidence that ST71 is derived from CC97 by insertion of a large fragment of DNA (approx. 412,500 bp) originating from a donor identical or closely related to ST3042. The only major difference is the absence of *cna* and a few surrounding genes in ST3042. However, there is only one known genome sequence of ST3042, the one described herein, and it might be assumed that different strains or sub-lineages of a putative “CC3042” might differ in the extent of the deletion, with some lacking *ica* and *his* operons only, while others lack *ica*, *his* and *cna*. Interestingly, there are two different variants of ST71 that differ in the size and in the exact localisation of the recombination breakpoints. Given the differences between Sardinian and all other ST71 sequences (including sequences in the cgMLST database from three Italian ST71 isolates), it is reasonable to assume that ST71 emerged twice by integration of ST3042 DNA into a CC97 backbone genome.

CC97 is a globally disseminated lineage that can be found in cattle but also in small ruminants, cervids, pigs, humans and in scavenging birds (which might have been exposed to ruminant carcasses) [3, 47–58]. Essentially nothing is known about ST3042, neither its geographic distribution nor its natural history including host species. The isolate sequenced herein originated from Italian cattle. There is only one entry in the MLST database (https://pubmlst.org/bigsdb?db=pubmlst_saureus_isolates&page=info&id=5580; accessed 2025, Dec. 8th and 2026, March 10th) and that isolate also originated from the Italian mainland. Unfortunately, there was no whole genome sequence provided, and the isolate was not available for sequencing. It is currently not known whether this lineage is restricted to Italy, and its natural host(s) remains elusive. It could be an extremely rare lineage in Mediterranean cattle. However, given the poor knowledge of methicillin-susceptible *S. aureus* in animals, it could also be a common strain in some wild animal species, possibly outside of Europe that just accidentally and sporadically got introduced into the cattle population. It might be argued that the ONT sequence described herein might be affected by sequencing errors, and that conclusions based on this sequence thus might be wrong or misleading. However, for three ST71 isolates, Illumina and ONT sequences were considered in parallel (see Supplemental file 2) and the degree of concordance was high indicating that ONT technology was sufficiently

accurate. Furthermore, not only the identity of certain alleles, but also the number of genes considered as well as the mere presence or absence of genes, especially within GIs, corroborate the conclusions. However, it would be interesting to analyse more genome sequences of ST3042, or closely related STs once they become available.

It also can be speculated that most European, including Irish and British ST71 form one clade originating back to one recombination event, while a second clade resulting from another recombination might be restricted to Sardinia, where this event might have taken place. This observation warrants further work on Sardinian ruminant strains of both CC97 and ST71. Sardinian ST71 also differ from other ST71 in the orientation of a large part of the genome. As the inversion affected the ST3042-derived region, this event must have taken place later in time than the recombination of CC97 and ST3042.

The acquisitions of the aureocin locus and of the pathogenicity island carrying *bap* by some isolates emphasise the ongoing evolution of this chimeric lineage as they further contribute to fitness in the competition with other bacteria and with host defences.

A significant question is how to reconcile the ST3042 hypothesis proposed here with a concept published earlier that suggested that ST71 was a chimera consisting of a large number of small genomic DNA fragments of different origins [32]. The two proposals are not necessarily contradictory. We argue that ST3042 (or a lineage closely related to ST3042) was an *immediate* predecessor to ST71 and donor of the part of the ST71 genome that does not originate from CC97. That does not rule out that ST3042 itself might be of chimeric/mosaic nature. For instance, it might have acquired the enterotoxin homologue ORF CM14 by horizontal gene transfer from one of the few lineages that harbour ORF CM14 (e.g., CC93, CC121, CC705, CC707 and CC772) into a still-to-be-determined parent strain. The ORF CM14-carrying genomic islands of CC707 (GenBank CP029751) and CC121 (CP026067, CP022717, CP022682, CP170404.1) have, compared to ST3042, identities of 99.90% and at least 99.77%, respectively. It theoretically could even be possible that ST71 and ST3042 are both chimeric lineages that carry a fragment of the genome of the same unknown donor lineage. With an increasing availability of genomes of *S. aureus* from animal sources and more diverse geographic regions, the natural history of ST71 and ST3042 might be more clearly resolved in future.

The fact that all lineages of *S. aureus*, however genetically diverse they might be, share just four *agr* alleles and two capsule types might indicate very far-reaching intermixing by a large number of past horizontal gene transfers. Thus, we assume that more lineages than previously anticipated (see Introduction) could be regarded as chimeric comprising fragments of two or even of multiple

parent strains. Of course, it becomes very complicated to analyse the provenience of regions in the genome when discussing series of several consecutive horizontal gene transfers, simply because at a certain point, distinctions between donors and recipients become blurred. In addition, the provenience of isolated genes or gene clusters consisting of only a few genes might be obscured by mutation and reversion. If two lineages shared the same cgMLST allele at one or two positions, this might be evidence of recombination, but it also could be a result of accidental mutations. If *several hundred* consecutive genes had identical cgMLST alleles, recombination is more likely than random mutation. Thus, it is easier to spot the provenience of large multi-gene segments and/or to attribute them to recombination as was indeed the case for ST71 as for any other previously described case of chimerism.

Finally, one might ask why a couple of recombination events in an obscure lineage of animal staphylococci are interesting or relevant. They are, because they shed light on the evolution of *S. aureus*, which is arguably one of the most versatile pathogens. Recombination might contribute to this success. Arguably, at least one of the recombinant ST71 clones apparently surpassed the parental lineage ST3042 in terms of abundance and geographic distribution, or at least were able to colonise European livestock, so that, from the viewpoint of a “Selfish Gene” [59], the “alliance” with genes from the pandemic CC97 lineage was a major success. The observation that ST71 emerged twice adds to the body of evidence that recombination is another, highly effective, mode of horizontal gene transfer and that it is no extraordinarily rare event in the evolutionary history of *S. aureus* but rather some kind of common process, likely to contribute to the extraordinary success of this pathogen.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-026-12777-w>.

Supplementary Material 1: Supplemental file 1a: Genome sequences of study strains. Supplemental file 1b: Statistics regarding ONT sequencing of the study strains. Supplemental file 2: cgMLST data and genomic islands. Unabridged version of Table 3. Supplemental file 3: Alignment of the genes around the first recombination breakpoint. Supplemental file 4: Alignment of the genes around the second recombination breakpoint.

Acknowledgements

We thank Dr Eoin Ryan, University College Dublin, School of Veterinary Medicine, Dublin, Ireland, for providing isolate FD516C. We thank Bernadette Hickey and Regina Heslin, Dairy Science Laboratory, Department of Agriculture, Food and the Marine, Backweston Campus, County Kildare, Ireland, for providing isolate 12_306.

Authors' contributions

DCC, HHE, OMK, AAM and RP identified and provided isolates used for the present study; SDB, CD, EM, and MR performed experiments and the initial analyses; SM analysed the sequences, created figures and wrote the draft

manuscript; CD wrote a part of the Materials & Methods section; DCC, OMK and RE reviewed the draft, SDB, SM, DCC, OMK, RP and RE acquired funding. All authors read and approved the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The Jena group was supported by a grant of the German Federal Ministry of Research, Technology and Space for the development of novel technologies for rapid diagnostic assays (LPI-BT5, grant number 13N15717) and by the Leibniz Society (project AMPeL, grant number AZ K548/2023). D.C.C. was supported by a grant from the Health Research Board (HRA-POR-2015-1051). O.M.K. was supported by funding from the MediCow project (RMIS1728). The Milan group was supported by a grant of the Filiera A.Q. Srl for an epidemiological study of intramammary infections in Italian bovine herds and the development of a control program.

Data availability

All relevant information, including genome sequences, is present in the manuscript and in the Supplemental files. Sequences of strains have been submitted to GenBank, accession numbers pending. The BioSample accession numbers are SAMN53808998, SAMN53808999, SAMN53809000, SAMN53809001, SAMN53809002, SAMN53809003, SAMN53809004, SAMN53809005, SAMN53809006, and SAMN53809007.

Declarations

Ethics approval and consent to participate

Not applicable. No isolates were purposefully collected for this study, but they were preserved specimens from diagnostic work and previous studies (see Table 1).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 December 2025 / Accepted: 18 March 2026

Published online: 18 April 2026

References

- van Belkum A, Melles DC, Nouwen J, van Leeuwen WB, van Wamel W, Vos MC, Wertheim HF, Verbrugh HA. Co-evolutionary aspects of human colonisation and infection by *Staphylococcus aureus*. *Infect Genet Evol*. 2009;9(1):32–47. <https://doi.org/10.1016/j.meegid.2008.09.012>.
- Heaton CJ, Gerbig GR, Sensus LD, Patel V, Smith TC. *Staphylococcus aureus* epidemiology in wildlife: a systematic review. *Antibiot (Basel)*. 2020;9(2):89. <https://doi.org/10.3390/antibiotics9020089>.
- Monecke S, Gavier-Widén D, Hotzel H, Peters M, Guenther S, Lazaris A, Loncaric I, Müller E, Reissig A, Ruppelt-Lorz A, et al. Diversity of *Staphylococcus aureus* isolates in European wildlife. *PLoS ONE*. 2016;11(12):e0168433.
- Xia G, Wolz C. Phages of *Staphylococcus aureus* and their impact on host evolution. *Infect Genet Evol*. 2014;21:593–601.
- Deghorain M, Van Melderden L. The Staphylococci phages family: an overview. *Viruses*. 2012;4(12):3316–35.
- Boakes E, Kearns AM, Ganner M, Perry C, Hill RL, Ellington MJ. Distinct bacteriophages encoding Panton-Valentine leukocidin (PVL) among international methicillin-resistant *Staphylococcus aureus* clones harboring PVL. *J Clin Microbiol*. 2011;49(2):684–92.
- Yamada T, Tochimaru N, Nakasuiji S, Hata E, Kobayashi H, Eguchi M, Kaneko J, Kamio Y, Kaidoh T, Takeuchi S. Leukotoxin family genes in *Staphylococcus aureus* isolated from domestic animals and prevalence of *lukM-lukF-PV* genes by bacteriophages in bovine isolates. *Vet Microbiol*. 2005;110(1–2):97–103.
- van Wamel WJ, Rooijackers SH, Ruyken M, van Kessel KP, van Strijp JA. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J Bacteriol*. 2006;188(4):1310–5.
- Coleman DC, Sullivan DJ, Russell RJ, Arbutnot JP, Carey BF, Pomeroy HM. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic

- conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. *J Gen Microbiol.* 1989;135(6):1679–97.
10. Azarian T, Cella E, Baines SL, Shumaker MJ, Samel C, Jubair M, Pegues DA, David MZ. Genomic epidemiology and global population structure of exfoliative toxin A-producing *Staphylococcus aureus* strains associated with staphylococcal scalded skin syndrome. *Front Microbiol.* 2021;12:663831.
 11. IWG-SCC. Classification of staphylococcal cassette chromosome *mec* (SCC-*mec*): guidelines for reporting novel SCC-*mec* elements. *Antimicrob Agents Chemother.* 2009;53(12):4961–7.
 12. Ito T, Katayama Y, Asada K, Mori N, Tsutsumimoto K, Tiensasitorn C, Hiramatsu K. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother.* 2001;45(5):1323–36.
 13. Ito T, Ma XX, Takeuchi F, Okuma K, Yuzawa H, Hiramatsu K. Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. *Antimicrob Agents Chemother.* 2004;48(7):2637–51.
 14. Shore AC, Coleman DC. Staphylococcal cassette chromosome *mec*: recent advances and new insights. *Int J Med Microbiol.* 2013;303(6–7):350–9.
 15. Shore AC, Deasy EC, Slickers P, Brennan G, O'Connell B, Monecke S, Ehrlich R, Coleman DC. Detection of staphylococcal cassette chromosome *mec* type XI carrying highly divergent *mecA*, *mecI*, *mecR1*, *blaZ*, and *ccr* genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother.* 2011;55(8):3765–73.
 16. Baines SL, Howden BP, Heffernan H, Stinear TP, Carter GP, Seemann T, Kwong J, Ritchie SR, Williamson DA. Rapid emergence and evolution of *Staphylococcus aureus* clones harbouring *fusC*-containing staphylococcal cassette chromosome elements. *Antimicrob Agents Chemother.* 2016;60(4):2359–65. <https://doi.org/10.1128/AAC.03020-15>
 17. Monecke S, Müller E, Braun SD, Armengol-Porta M, Bes M, Boswihi S, El-Ashker M, Engelmann I, Gawlik D, Gwida M, et al. Characterisation of *S. aureus*/MRSA CC1153 and review of mobile genetic elements carrying the fusidic acid resistance gene *fusC*. *Sci Rep.* 2021;11(1):8128.
 18. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol.* 2000;38(3):1008–15.
 19. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 2018;3:124.
 20. Robinson DA, Enright MC. Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol.* 2004;186(4):1060–4.
 21. Thomas JC, Godfrey PA, Feldgarden M, Robinson DA. Draft genome sequences of *Staphylococcus aureus* sequence type 34 (ST34) and ST42 hybrids. *J Bacteriol.* 2012;194(10):2740–1.
 22. Gawlik D, Ruppelt-Lorz A, Müller E, Reissig A, Hotzel H, Braun SD, Soderquist B, Ziegler-Cordts A, Stein C, Pletz MW, et al. Molecular investigations on a chimeric strain of *Staphylococcus aureus* sequence type 80. *PLoS ONE.* 2020;15(10):e0232071.
 23. Monecke S, Braun SD, Collatz M, Diezel C, Müller E, Reinicke M, Cabal Rosel A, Feßler AT, Hanke D, Loncaric I, et al. Molecular characterization of chimeric *Staphylococcus aureus* strains from waterfowl. *Microorganisms.* 2024;12(1):96. <https://doi.org/10.3390/microorganisms12010096>.
 24. Fetsch A, Kraushaar B, Kasbohrer A, Hammerl JA. Turkey meat as source of CC9/CC398 methicillin-resistant *Staphylococcus aureus* in humans? *Clin Infect Dis.* 2017;64(1):102–3.
 25. Strommenger B, Bräulke C, Heuck D, Schmidt C, Pasemann B, Nübel U, Witte W. *spa* typing of *Staphylococcus aureus* as a frontline tool in epidemiological typing. *J Clin Microbiol.* 2008;46(2):574–81.
 26. Monecke S, Coombs G, Shore AC, Coleman DC, Akpaka P, Borg M, Chow H, Ip M, Jatzwaik L, Jonas D, et al. A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. *PLoS ONE.* 2011;6(4):e17936.
 27. Nimmo GR, Steen JA, Monecke S, Ehrlich R, Slickers P, Thomas JC, Appleton S, Goering RV, Robinson DA, Coombs GW. ST2249-MRSA-III: a second major recombinant methicillin-resistant *Staphylococcus aureus* clone causing healthcare infection in the 1970s. *Clin Microbiol Infect.* 2015;21(5):444–50.
 28. Burgold-Voigt S, Monecke S, Simbeck A, Holzmann T, Kieninger B, Liebler-Tenorio EM, Braun SD, Collatz M, Diezel C, Müller E et al. Characterisation and molecular analysis of an unusual chimeric methicillin resistant *Staphylococcus aureus* strain and its bacteriophages. *Front Genet.* 2021;12(1823). <https://doi.org/10.3389/fgene.2021.723958>.
 29. Parisi A, Caruso M, Normanno G, Latorre L, Sottili R, Miccolupo A, Fraccalvieri R, Santagada G. Prevalence, antimicrobial susceptibility and molecular typing of methicillin-resistant *Staphylococcus aureus* (MRSA) in bulk tank milk from southern Italy. *Food Microbiol.* 2016;58:36–42.
 30. Budd KE, McCoy F, Monecke S, Cormican P, Mitchell J, Keane OM. Extensive genomic diversity among bovine-adapted *Staphylococcus aureus*: evidence for a genomic rearrangement within CC97. *PLoS ONE.* 2015;10(8):e0134592.
 31. Cormican P, Keane OM. Complete genome sequences of sequence type 71 (ST71) and ST97 *Staphylococcus aureus* isolates from bovine milk. *Microbiol Resour Announc.* 2018;7(5):e00954–18. <https://doi.org/10.1128/MRA.00954-18>.
 32. Spoor LE, Richardson E, Richards AC, Wilson GJ, Mendonca C, Gupta RK, McAdam PR, Nutbeam-Tuffs S, Black NS. Recombination-mediated remodeling of host–pathogen interactions during *Staphylococcus aureus* niche adaptation. *Microb Genomics.* 2015;1(4):e000036. <https://doi.org/10.1099/mgen.0.000036>.
 33. Dicks J, Turnbull JD, Russell J, Parkhill J, Alexander S. Genome sequencing of a historic *Staphylococcus aureus* collection reveals new enterotoxin genes and sheds light on the evolution and genomic organization of this key virulence gene family. *J Bacteriol.* 2021;203(10):e00587–20. <https://doi.org/10.1128/JB.00587-20>.
 34. Moawad AA, Dore S, Lollai S, El-Adawy H, Brangsch H, Monecke S, Ehrlich R, Braun SD, Neubauer H, Tomaso H. Genomic characterization of *Staphylococcus aureus* isolated from mastitis in small ruminants in Sardinia, Italy. *Front Microbiol.* 2025;16:1661122. <https://doi.org/10.3389/fmicb.2025.1661122>.
 35. Kjems E. Two Variants of *Staphylococcus aureus* WOOD 46 (NCTC 7121) differing in respect to alpha toxin production. *J Bacteriol.* 1963;86(5):1127–8.
 36. Poshvina DV, Dilbaryan DS, Kasyanov SP, Sadykova VS, Lapchinskaya OA, Rogozhin EA, Vasilchenko AS. *Staphylococcus aureus* is able to generate resistance to novel lipopeptide antibiotic gausemycin A. *Front Microbiol.* 2022;13:963979.
 37. Magro G, Biffani S, Minozzi G, Ehrlich R, Monecke S, Luini M, Piccinini R. Virulence genes of *S. aureus* from dairy cow mastitis and contagiousness risk. *Toxins.* 2017;9(6).
 38. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
 39. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017;27(5):737–46.
 40. Luan T, Commichaux S, Hoffmann M, Jayeola V, Jang JH, Pop M, Rand H, Luo Y. Benchmarking short and long read polishing tools for nanopore assemblies: achieving near-perfect genomes for outbreak isolates. *BMC Genomics.* 2024;25(1):679.
 41. Watanabe S, Ito T, Sasaki T, Li S, Uchiyama I, Kishii K, Kikuchi K, Skov RL, Hiramatsu K. Genetic diversity of staphylocoagulase genes (*coa*): insight into the evolution of variable chromosomal virulence factors in *Staphylococcus aureus*. *PLoS ONE.* 2009;4(5):e5714.
 42. Wan TW, Khokhlova OE, Iwao Y, Higuchi W, Hung WC, Reva IV, Singur OA, Gostev VV, Sidorenko SV, Peryanova OV, et al. Complete circular genome sequence of successful ST8/SCC-*mecIV* community-associated methicillin-resistant *Staphylococcus aureus* (OC8) in Russia: one-megabase genomic inversion, IS256's spread, and evolution of Russia ST8-IV. *PLoS ONE.* 2016;11(10):e0164168.
 43. Monecke S, Burgold-Voigt S, Braun SD, Diezel C, Liebler-Tenorio EM, Müller E, Nassar R, Reinicke M, Reissig A, Senok A et al. Characterisation of PVL-Positive *Staphylococcus argenteus* from the United Arab Emirates. *Antibiot (Basel).* 2024;13(5):401. <https://doi.org/10.3390/antibiotics13050401>.
 44. Netz DJ, Sahl HG, Marcelino R, dos Santos Nascimento J, de Oliveira SS, Soares MB, do Carmo de Freire Bastos M. Molecular characterisation of aureocin A70, a multi-peptide bacteriocin isolated from *Staphylococcus aureus*. *J Mol Biol.* 2001;311(5):939–49.
 45. Cucarella C, Solano C, Valle J, Amorena B, Lasa I, Penadés JR. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *J Bacteriol.* 2001;183(9):2888–96.
 46. Snel GG, Monecke S, Ehrlich R, Piccinini R. Molecular characteristics of bap-positive *Staphylococcus aureus* strains from dairy cow mastitis. *J Dairy Res.* 2015;82(3):312–6.
 47. Menegotto F, Gonzalez-Cabrero S, Cubero A, Cuervo W, Munoz M, Gutierrez MP, Simarro M, Bratos MA, Orduna A. Clonal nature and diversity of resistance, toxins and adhesins genes of methicillin-resistant *Staphylococcus aureus* collected in a Spanish hospital. *Infect Genet Evol.* 2012;12(8):1751–8.
 48. Keinprecht H, Irimaso E, Rosel AC, Stessl B, Ntakirutimana C, Marek L, Fischer OW, Szostak MP, Zochbauer J, Wittek T, et al. Diversity of *Staphylococcus aureus* associated with mastitis from dairy cows in Rwanda. *J Glob Antimicrob Resist.* 2024;36:326–35.

49. Gomez-Sanz E, Torres C, Lozano C, Fernandez-Perez R, Aspiroz C, Ruiz-Larrea F, Zarazaga M. Detection, molecular characterization, and clonal diversity of methicillin-resistant *Staphylococcus aureus* CC398 and CC97 in Spanish slaughter pigs of different age groups. *Foodborne Pathog Dis*. 2010;7(10):1269–77.
50. Ruiz-Ripa L, Gómez P, Alonso CA, Camacho MC, de la Puente J, Fernández-Fernández R, Ramiro Y, Quevedo MA, Blanco JM, Zarazaga M, et al. Detection of MRSA of Lineages CC130-*mecC* and CC398-*mecA* and *Staphylococcus delphini-Inu(A)* in Magpies and Cinereous Vultures in Spain. *Microb Ecol*. 2019;78(2):409–15.
51. Luzzago C, Locatelli C, Franco A, Scaccabarozzi L, Gualdi V, Viganò R, Sironi G, Besozzi M, Castiglioni B, Lanfranchi P, et al. Clonal diversity, virulence-associated genes and antimicrobial resistance profile of *Staphylococcus aureus* isolates from nasal cavities and soft tissue infections in wild ruminants in Italian Alps. *Vet Microbiol*. 2014;170(1–2):157–61.
52. Loncaric I, Keinprecht H, Irimaso E, Cabal-Rosel A, Stessl B, Ntakirutimana C, Marek L, Fischer OW, Szostak MP, Oberrauch C et al. Diversity of *Staphylococcus aureus* isolated from nares of ruminants. *J Appl Microbiol*. 2024;135(1):lxad304. <https://doi.org/10.1093/jambio/lxad304>.
53. Senok A, Nassar R, Kaklamanos EG, Belhouli K, Abu Fanas S, Nassar M, Azar AJ, Muller E, Reissig A, Gawlik D, et al. Molecular characterization of *Staphylococcus aureus* isolates associated with nasal colonization and environmental contamination in academic dental clinics. *Microb Drug Resist*. 2020;26(6):661–9.
54. Senok A, Ehricht R, Monecke S, Al-Saedan R, Somily A. Molecular characterization of methicillin-resistant *Staphylococcus aureus* in nosocomial infections in a tertiary-care facility: emergence of new clonal complexes in Saudi Arabia. *New Microbes New Infect*. 2016;14:13–8.
55. Monecke S, Skakni L, Hasan R, Ruppelt A, Ghazal SS, Hakawi A, Slickers P, Ehricht R. Characterisation of MRSA strains isolated from patients in a hospital in Riyadh, Kingdom of Saudi Arabia. *BMC Microbiol*. 2012;12(1):146.
56. Kechrid A, Perez-Vazquez M, Smaoui H, Hariga D, Rodriguez-Banos M, Vindel A, Baquero F, Canton R, Del Campo R. Molecular analysis of community-acquired methicillin-susceptible and resistant *Staphylococcus aureus* isolates recovered from bacteraemic and osteomyelitis infections in children from Tunisia. *Clin Microbiol Infect*. 2011;17(7):1020–6.
57. Gostev V, Kruglov A, Kalinogorskaya O, Dmitrenko O, Khokhlova O, Yamamoto T, Lobzin Y, Ryabchenko I, Sidorenko S. Molecular epidemiology and antibiotic resistance of methicillin-resistant *Staphylococcus aureus* circulating in the Russian Federation. *Infect Genet Evol*. 2017;53:189–94.
58. Achek R, El-Adawy H, Hotzel H, Tomaso H, Ehricht R, Hamdi TM, Azzi O, Monecke S. Short communication: Diversity of staphylococci isolated from sheep mastitis in northern Algeria. *J Dairy Sci*. 2020;103(1):890–7.
59. Dawkins R. *Das egoistische Gen (The selfish gene)*. German ed. Heidelberg: Spektrum; 1994.

Publisher's note

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