Origin and evolutionary dynamics of multi-drug resistant and highly virulent community-associated methicillin-resistant *Staphylococcus aureus* ST772-SCCmec V lineage

¹*Yamuna Devi Bakthavatchalam, ¹Karthick Vasudevan, ²ShomaVinay Rao, ³Santosh Varughese, ⁴Priscilla Rupali, ⁵Maki Gina, ⁵Marcus Zervos, ⁶John Victor Peter, ¹Balaji Veeraraghavan*

¹Department of Clinical Microbiology, Christian Medical College, Vellore-632004

²Department of Surgical Intensive Care Unit (SICU), Christian Medical College, Vellore-632004

³Department of Nephrology, Christian Medical College, Vellore-632004

⁴Division of Infectious Disease, Christian Medical College, Vellore-632004

⁵Division of Infectious Disease, Henry Ford Health System, Detroit, Michigan - 48202

⁶Department of Critical Care Medicine, Christian Medical College, Vellore-632004

Corresponding author:

Dr Balaji Veeraraghavan
Professor
Department of Clinical Microbiology
Christian Medical College
Vellore – 632004
E-mail ID: vbalaaji@cmcvellore.ac.in
Contact no: +91- 9442210555

Running title: ST772 *S. aureus* genomic investigation

25

26 **Structured synopsis (245/250):**

27 **Background:**

28 Community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA) are
29 increasing in prevalence across the world. However, studies on the molecular
30 epidemiology and the genomic investigation of MRSA are limited in India.

31 **Objectives:**

32 To understand the molecular epidemiology of MRSA and to reconstruct the origin and
33 evolution of *S. aureus* belonged to the sequence type (ST772).

34 **Methods:**

35 A total of 233 non-repetitive MRSA isolates were screened for the presence
36 staphylococcal cassette chromosome (*SCCmec*) types, multi-locus sequence types
37 (MLST) and staphylococcal protein A (*spa*) types. Whole genome sequence data of
38 ST772-*SCCmec* V (n=32) isolates were generated and analysed along with the publically
39 available ST772-*SCCmec* V (n=273) genome.

40 **Results:**

41 ST772 (27%), ST22 (19%) and ST239 (16%) were found as the predominant STs.
42 Analysis of the core SNPs using Bayesian time scaled phylogenetic analysis showed
43 ST772-*SCCmec* V was emerged on the Indian subcontinent in 1960s. The acquisition of
44 integrated resistance plasmid (IRP) in the ST772-*SCCmec* V lineage during 1990s,
45 fixation of *SCCmec* V (5C2) and the double serine mutations (S84L, S80Y) appears to
46 have played a key role in the successful expansion. The IRP carries the loci for multiple

antibiotic resistant genes: beta-lactam (*blaZ*), aminoglycosides (*aphA3-sat-aadE*),
macrolide (*mphC*), macrolide-lincosamide-streptogramin B (*msrA*) and bacitracin (*bacA*,
bacB).

Conclusion:

The Panton Valentine Leukocidin (PVL) positive ST772 and ST22 MRSA lineages are
observed in the hospital settings. ST772-SCC*mec* V has the multi-drug resistance trait of
hospital-associated (HA) MRSA and the epidemiological characteristics of CA-MRSA.
The antimicrobial use pattern may have driven the spread and survival of ST772 MRSA
in hospitals.

67

68 **Introduction:**

69 Methicillin-resistant *Staphylococcus aureus* (MRSA) is of significant clinical concern in
70 causing both community-associated (CA) and healthcare-associated (HA) infections.

71 MRSA was first described in 1996.¹ It has merged through the acquisition of different
72 staphylococcal cassette chromosome *mec* (SCC *mec*) carrying methicillin resistance
73 determinant *mecA*.² The HA-MRSA strains carries the SCC*mec* types I, II or III and are
74 often multi-drug resistant. Typically, the CA-MRSA strains carries the SCC*mec* types IV
75 or V, a Panton-Valentine leukocidin (PVL) toxin and are frequently susceptible to non-
76 beta-lactam antibiotics.³

77 Most of the HA-MRSA clones belong to the clonal complex (CC), CC5, CC8, CC22, CC30
78 and CC45.⁴ In contrast, CA-MRSA clones are geographically more diverse. In the Asia-
79 pacific region, five distinct sequence types (STs); ST59, ST30, ST72, ST8, and ST772 are
80 the predominant CA-MRSA clones and have often spread to hospitals.⁵ An increasing
81 resistance to non-beta-lactam antibiotics have been reported in CA-MRSA.⁶ In India,
82 MRSA accounts for about 37% of *S. aureus* infections.^{7,8} However, there are only limited
83 studies on the molecular epidemiology of *S. aureus*.

84 In early 2000, most of the Indian MRSA isolates belonged to the ST239-SCC*mec* III
85 lineage and were restricted to hospitals.⁹⁻¹² In 2004, a novel CA *S. aureus* clone, ST772-
86 Methicillin-susceptible *S. aureus* (MSSA) isolated from hospitals in Bangladesh and
87 ST772-SCC*mec* V MRSA (Bengal bay clone) isolated from the community and hospitals
88 in India has been described.^{12,13} During the same year, a variant of epidemic MRSA
89 (EMRSA-15) carries PVL, SCC*mec* IV and belongs to ST22 has also been reported in

India.¹² After 2004, ST772 and ST22 has increased in prevalence to become the dominant STs in both community and hospital infections in India.¹⁴⁻¹⁹ In the last decade, ST772-SCC*mec* V has proven to be successful at disseminating into hospitals, begin with localized expansion and rapidly progressing to an intercontinental spread.^{20,21}

Genomic studies have brought both understanding about the evolution of *S. aureus* and new insights into the genetic determinants contributing to its spread and survival. The well-studied HA-MRSA genomes are mostly restricted to the Hungarian clone (ST239-SCC*mec* III) and EMRSA-15 (ST22-SCC*mec* IV).^{22,23} Other well-studied CA-MRSA genomes are the USA300 clone (ST8-SCC*mec* IV) and a live-stock associated MRSA, the ST398-SCC*mec* V lineage.^{24, 25}

The present study objectives were i) to provide preliminary data for isolate selection that belongs to ST772-SCC*mec* V lineage for further genomic investigation, 233 MRSA isolates were screened for genotypes using SCC*mec* typing, multi-locus sequence typing (MLST) and staphylococcal protein A (*spa*) typing, ii) to reconstruct the origin of ST772-SCC*mec* V isolates collected between 2004 and 2018, iii) to establish a high resolution phylogeny and evolutionary analysis of this emerging lineage, iv) to trace the genetic changes associated with its evolution and success and v) to characterise the pangenome of the ST772-SCC*mec* V lineage.

Materials and methods:

Bacterial isolates:

Non-repetitive MRSA (n=233) isolates (one isolate per patient) recovered from the blood culture samples during 2013-2018 were included in this study. MRSA isolates were identified using standard bacteriological methods including gram staining, culture

and tube coagulase test.²⁶ The study was conducted in a 2600 bedded tertiary care, Christian Medical College and hospital, Vellore, India and approved by the Institutional Review Board of Christian Medical College, Vellore (IRB.Min.no. 10643, dated on April, 2017).

Molecular characterisation of MRSA:

SCCmec types were determined using multiplex PCR as described by Zhang *et al.*²⁷ The subtypes of SCCmec IV were identified as described previously.²⁸ MLST of *S. aureus* was performed as defined by Enright *et al.*²⁹ STs and CC were assigned according to the MLST database (<https://pubmlst.org/saureus/>). *spa* typing was performed as described by Harmsen *et al.*³⁰ Assignment of the *spa* types were carried out using the public *spa* type database Ridom SpaServer (www.spaserver.ridom.de).

Genome sequencing and assembly:

A representative subset of 32 isolates of the predominant ST772-SCCmec V lineage were further analysed using whole genome sequencing and compared with the available published genomes. Then, the genetic lineage, virulence and resistant determinants were described.

Ion Torrent sequencing:

Whole genome sequencing was performed using IonTorrent PGM platform (Life Technologies, Carlsbad, CA). The sequencing library was prepared using 200ng of genomic DNA which was sheared to a fragment length of 400bp. Library preparation and sequencing were performed according to the manufacturer's instruction using Ion Plus Fragment Library Kit (Life Technologies, California, United States). De novo

assembly of the raw reads generated from ion torrent were assembled using
AssemblerSPAdes software v4.4.0.1 in Torrent suite server version 4.4.3.

Genome annotation and downstream analysis:

The genome was annotated using the RAST server (Rapid Annotation using Subsystem
Technology, <http://rast.nmpdr.org>)³¹ available from the PATRIC database
(Pathosystems Resource Integration Centre, <https://www.patricbrc.org>)³² and the
National Center for Biotechnology Information (NCBI) Prokaryotic Genome Automatic
Annotation Pipeline ([www.ncbi.nlm.nih.gov/genomes/ static/Pipeline.html](http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html)). SCCmec
types were determined using the SCCmec finder database.³³ MLST of the isolates were
reconfirmed using the MLST finder (<https://cge.cbs.dtu.dk/services/MLST/>) which is
available from the Centre for Genomic Epidemiology database.

Genome analysis for virulence and resistant genes:

Virulence factors were identified using virulence factor database (VFDB)³⁴ and the
acquired virulence genes using VirulenceFinder
(<https://cge.cbs.dtu.dk/services/VirulenceFinder/>). Acquired antimicrobial resistant
(AMR) genes were identified by using the following different AMR databases; BLAST
search against comprehensive antibiotic resistance database (CARD)⁵, ResFinder 3.2
(<https://cge.cbs.dtu.dk/services/ResFinder/>), KmerResistance 3.1
(<https://cge.cbs.dtu.dk/services/KmerFinder/>), Mykrobe predictor³⁶ and
Pathogenwatch (<https://www.sanger.ac.uk/science/tools/pathogenwatch>). Insertion of
the phage genome was identified using PHASTER.³⁷ Pathogenicity Islands were
identified using the Pathogenicity island database.³⁸

Genome comparison:

The original ST772-SCCmec V global data of Steinig *et al*³⁹, Prabhakara *et al*⁴⁰ and Balakuntla *et al*⁴¹(Supplementary Table 1) were supplemented with the genome of an additional 32 isolates sequenced in this study collected during 2013-2018. Sequence reads of *S. aureus* ST772 isolates (n=273) were retrieved from the NCBI and European Nucleotide Archive (ENA) database and assembled with SPAdes v.3.12⁴². DAR4145 (accession no. CP010526) deposited from Mumbai was used as the reference genome.

The capsular biosynthesis gene (*capA-G* and *capL-P*), an accessory gene regulator (*agrBDCA*), a DNA repair protein (*radC*), the type-I restriction-modification system (*hsdR*) and the type III restriction system were analysed for mutations. The effect of amino acid substitution on the protein function was predicted using the sorting intolerant from tolerant (SIFT) score.⁴³

Single nucleotide polymorphism (SNP) based phylogenetic analysis:

Phylogenetic analysis was performed against the publically available global genomes. The genomes were aligned to the reference genome using BWA mem algorithm.⁴⁴ Variants were called using Snippy v.4.5.1.⁴⁵ and FreeBayes⁴⁶ with a minimum base quality of 20, a minimum read coverage of 10X and a 90% concordance at a locus for a variant were selected. A whole SNP genome alignment and core SNPs genome alignment of all genomes was generated with snippy-core. Core SNPs were defined as being present in all isolates and were extracted using snippy-core with default settings was used to infer a phylogeny. The recombination free core genome SNP alignment file generated by Gubbins was used as the input to compute the mean evolutionary rate of the genomes and time of the most recent common ancestor (MRCA).⁴⁷ The RaXML-NG

0.5.0 was used to generate a maximum-likelihood (ML) tree based on 6344 variant sites in the core genome.⁴⁸ The general time reversible model (GTR + Γ) of nucleotide substitution was implemented with ascertainment bias correction (Lewis) and 100 bootstrap replicates as suggested by Steinig *et al.*³⁹ The tree with the highest likelihood was midpoint rooted and visualised with iTOL.⁴⁹ The study isolates and the clades were assigned to previously described lineages.³⁹ A hierarchical Bayesian analysis of population structure (hierBAPS) clustering model was used to support phylogenetic groupings by using iterative clustering to the depth of 10 levels and a pre-specified maximum of 20 clusters.^{50,51}

Bayesian time scaled phylogenetic analysis:

Temporal signal in the ML phylogeny for ST772 *S. aureus* was investigated using TempEST v 1.5.1 (<http://tree.bio.ed.ac.uk>). The relationship between root to tip distance and the time of isolation were analysed. Root-to-tip regressions (R2) may indicate the evolutionary rate and probable outliers among the isolates. The temporal phylogenetic structure was determined using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v1.10.4 package after the initial assessment of the temporal signal.⁵² Sources and date of isolation of 305 ST772 *S. aureus* isolates used for temporal and BEAST analysis are listed in Supplementary Table 5. Evolutionary rates and tree topologies were determined with the generalized time reversible (GTR) substitution models with gamma distributed among-site rate variation with four rate categories (GTR+ Γ 4+I). The strict (assumes that branches in the tree evolves at constant evolution rate) and uncorrelated relaxed molecular clocks (assumes that each branch in the tree have independent evolution rate) were implemented with three demographic models, Constant size, Exponential and Bayesian skyline coalescent tree priors. The MCMC

chain was run for 500 million steps with sampling of 20000 generations. We performed an additional duplicate run of 500 million steps and combined the log using LogCombiner v1. 8.2 when the effective sample size (ESS) estimates did not reach 200.⁵³ The convergence and mixing of each run was manually inspected Tracer.v.1.7 to ensure that all the runs converged to an ESS of >200.⁵⁴ To determine the best-fitting coalescent model to describe changes in effective population size over time, log marginal likelihoods were calculated using path sampling and stepping stone sampling methods. Finally, Bayes factor was used to determine the best fit model⁵⁵, a burn in of 20% was discarded from the runs and the maximum clade credibility (MCC) tree was generated using Treeannotator v.1.8.2.⁵⁶ BEAST output was analyzed using Tracer v1.7, with uncertainty in parameter estimates reflected as the 95% highest probability density (HPD). The annotated phylogenetic tree was visualized using FigTree v.1.4.6 and the Bayesian skyline plot was reconstructed using Tracer.v.1.7.^{57,58}

Pangenome analysis

Pangenome analysis of ST772 *S. aureus* (n=305) was performed using Roary v. 3.11.2 with default settings.⁵⁹ The genomes were annotated using Prokka v.2.1.4.⁶⁰ The genes that were common in all compared strains (core genes) and accessory genes were extracted and was used to construct a phylogenetic tree. Core genes (99% ≤ strains ≤ 100%), soft core genes (95% ≤ strains < 99%), shell genes (15% ≤ strains < 95%), cloud genes (0% ≤ strains < 15%) and total genes (0% ≤ strains ≤ 100%) were calculated. In addition, the presence or absence matrix representing accessory genes were graphed and visualized using roary_plots.py python script (<https://github.com/sangerpathogens/Roary>). Mobile genetic elements present in the genome were identified using comparative genomic analysis and manual curation.

Accessory genome analysis was divided into *S. aureus* Pathogenicity Island (SaPI), prophage, plasmids, SCCmec elements and transposons that are known to contribute major source of variation between *S. aureus* genomes.²² Mobile genetic elements (MGEs) in the genomic sequences were analysed using comparative genomic analysis and manual curation. IRP, transposons and insertion sequence (IS) of DAR4145 were used as the reference sequences. In addition, IS mapper and was used for the screening of IS elements. Detailed comparisons of the genome sequences were performed on the de novo assemblies using BLASTN. Hypothetical proteins are those did not have any matches either with core or MGEs of reference genome sequence (DAR4145). The core genome elements were determined using Spine V0.3.1 with default cutoff (Minimum nucleotide percent identity of 85% and 100% minimum threshold to identify as core).⁶¹ Additionally, accessory genomic elements were identified using AGEnt v0.3.1 that were not identified as core genomes by Spine.⁶¹ Further, the accessory genomes from all the isolates were combined and functional annotation was determined using eggNOG-mapper.⁶²

Results:

Preliminary molecular characterisation of MRSA isolates:

SCCmec V was the predominant gene in 42% (n=98) of the isolates followed by SCCmec III in 27% (n=63), SCCmec IV in 13% (n=31), SCCmec I in 7% (n=16) and SCCmec II in 3% (n=7) of the isolates. The SCCmec IV subtypes were IVa (n=9), IVc (n=7), IVd (n=14), and IVh (n=1). Notably, eighteen isolates had multiple SCCmec types; fifteen isolates carried SCCmec III and SCCmec V and one isolate was identified with each of the following SCCmec I and SCCmec IVd; SCCmec IVd and SCCmec V; SCCmec IVa and SCCmec V.

MLST analysis revealed ten different CC and three singletons (ST616, ST1598, ST1947) (Fig 1). Among the CCs, five high-risk clones (HRC) including CC5, CC8, CC22, CC30, and CC45 were observed. Three major CCs; CC1 (ST1, ST772), CC22 (ST22, ST217, ST636, ST896, ST1037, ST2371, ST3976) and CC8 (ST8, ST239, ST368, ST630, ST1803, ST3324) accounted for 82% (n=192) of the isolates. Other minor CCs and their associated STs were given in Table S1. A total of 32 different STs were identified. The most common were ST772 in 27% (n=63), ST22 in 19% (n=44), ST239 in 16% (n=37) and ST368 in 9% (n=20) of the isolates (Fig 1). Although, the STs were observed to be diverse, most of them were recognised as a SLV or DLV of the dominant STs (Table S2). Thirty distinct *spa* types were identified. The *spa* type t657 was identified in 34% (n=73) of isolates followed by t425 in 18% (n=34), t037 in 12% (n=25), t030 in 8% (n=18) and t852 in 10% (n=17). The remaining 25 different *spa* types were found in fewer than five isolates (Fig 1, Table S3).

ST772-SCCmec V-t657 lineage was observed in 22% (n=52), ST239-SCCmec III lineage sharing the *spa* types t037 and t030 in 15% (n=35) and ST368-SCCmec III-t425 lineage in 8% (n=19) of the isolates were identified. In contrast, SCCmec and *spa* types were highly diverse among the ST22 MRSA strains (Fig 1, Table S3).

Genomic analysis of *S. aureus* ST772-SCCmec V:

Phylogenetic analysis:

A representative subset of 32 ST772-SCCmec V MRSA isolates characterized in the preliminary screening were included and compared with the genomes that are available in the public database. The country and the year of isolation of ST772 *S. aureus* isolates are given in the Supplementary Fig 1. ST772-SCCmec V genome data was comprised of

MRSA (n=270) and MSSA (n=34) isolates. Ten basal strains (4 MSSA, 6 MRSA) were also included. Mapping of ST772 *S. aureus* isolate against the reference genome resulted in an alignment of 6344 core SNPs, which were used for phylogenetic inference. A maximum likelihood phylogenetic tree generated from the core genome SNP is presented in Fig 2a. ST772-SCCmec V isolates were grouped into two distinct clades; clade I represented basal strains, the earliest was found to be MSSA which was isolated from India and Bangladesh in 2004. The clade II was recognised as a globally distributed monophyletic group and diverged into three different subclades (A, B, C). Majority of the isolates clustered in clade IIB (56%, n=169) which were considerably more diverse and dominant than early branching clade IIA (24%, n=73) and emerging clade IIC (17%, n=52) isolates. The phylogenetic comparison of Indian ST772 *S. aureus* isolates against isolates available from other countries showed that Indian isolates were distributed across all the subclades and also the clade containing basal strains.

Population structure analysis using RhierBAPS revealed two distinct BAPS groups (Fig 2b). The basal strains and isolates of subclade IIA were classified as BAPS population cluster 1, while isolates of subclade II B and C were categorised as BAPS population cluster 2. In BAPS population cluster 1, sequence comparison of 83 isolates (basal strains and subclade IIA) against the reference genome DAR4145 identified 3306 core SNPs (among a core genome of 2.8 Mbp). Similarly, in BAPS population cluster 2, a total of 3678 core SNPs (among a core genome of 2.8 Mbp) were identified by aligning the sequences of 221 isolates (subclade IIB and C) against the reference genome.

Bayesian time scaled phylogenetic analysis:

As there was a strong temporal signature for the observed mutations in the global ST772 *S. aureus* population, we explored the temporal structure of ST772 *S. aureus*

isolates using Bayesian phylogenetic methods. The root to tip analysis revealed a strong correlation (correlation coefficient 0.6862 and R squared 0.4709) between the time of isolation and distance from root suggesting the temporal clock like evolution in the lineage (Fig 3). A maximum clade credibility tree was reconstructed based on the concatenated core SNPs and is shown in Fig 4. The mean substitution rate of the ST772 *S. aureus* population was estimated to be 1.16×10^{-6} substitution per site per year. The time-scaled phylogenetic reconstruction suggest that the MRCA of ST772 *S. aureus* with initial divergence was estimated to be from the year 1959 (95% HPD, 1936-1979). This was followed by the emergence of dominant clade II isolates and its subgroups in 1987 (95% HPD, 1981-1995). Bayesian hierarchical clustering using core SNPs segregated *S. aureus* ST772 isolates that had IRP and a variant of SCCmec V (5C2) into separate subclades (Fig 4). This infers that the ST772 *S. aureus* resistant subgroups expanded rapidly after 1987 and has since sustained.

The generated BSP estimated the demographic changes of ST772 *S. aureus* genome (n=305) over time (Fig 5). The effective population size of ST772 was relatively stable from early the 1950s to the late 1970s. Notably, two mild expansion events were noticed. Initial event was associated with the acquisition of IRP, and SCCmec V (5C2) in 1987 followed by another event which replaced the double serine mutations *gyrA* (S84L) with *grlA* (S80F) by *gyrA* (S84L) with *grlA* (S80Y) in 1993. These genetic changes resulted in a sharp increase in ST772-SCCmec V effective population size in 1997. At this stage, the population size increased by one order of magnitude. After 2010, there was a sharp decline in the effective population size and recently there has been a noticeable plateau.

Virulence factors:

Virulence factors identified in ST772 *S. aureus* are given in the Table S3. Almost all of the isolates (99%, n=301) carried a 59.4 kb Φ IND772 prophage containing PVL-operon (*lukS* and *lukF*) and a staphylococcal enterotoxin A (*sea*). Three different genomic islands carrying various virulence factors were identified; a 41 kb, vSa-alpha genomic island containing an exotoxin cluster (*set6*, *set7*, *set8*, *set10*, *set11*, *set12*, *set13*, *set14*, *set15*); a 25.2 kb vSa-beta genomic island containing an enterotoxin gene cluster (*sec*, *seg*, *sei*, *sem*, *sen*, *seo*); and a 24.6 kb, vSa-gamma genomic island contains an extracellular fibrinogen binding protein (*ecb*), a fibrinogen binding protein (*fbp*), an alpha hemolysin (*hla*) and an exfoliative toxin A (*eta*) (Fig S2). All the three genomic islands were universally seen across the basal strains and isolates of globally disseminated clade II (subclades IIA, B and C) (Fig 6). In addition, all the isolates were found to carry *aur*, *scn*, *hld*, *ebs* and a biofilm encoding *ica* operon (*icaADBC*) in the genome (Fig 6 and Table S3).

Gene sequences of *capA-G*, *capL-P*, *agrBDCA*, *radC*, type I (*hsdR*) and type III restriction system were highly conserved among the ST772 lineage. The substitutions and premature stop codons identified in these genes are listed in the Table S4. The substitutions in *cap5B* (K55N, K55Q) and *cap5D*(K7E) were predicted to affect the protein function with a SIFT score of 0.001.

SCCmec types:

A 47.2 kb composite SCCmec V (5C2 and 5) and a 29.2kb SCCmec V (5C2) variant were recognised in 7% (n=21) and 76% (n=230) of isolates respectively (Fig 2a). All the basal MRSA (n=6) isolates harboured a large composite SCCmec V (5C2 and 5) element.

The diversity of MRSA carrying *SCCmec* types decreased as ST772 clade II diverged into subgroups: clade IIA comprised of MSSA (n=28), *SCCmec* V (5C2 and 5, n=17) and *SCCmec* V (5C2, n=26); in contrast subclades IIB and IIC were exclusively carried *SCCmec* V (5C2) (Fig 2a). A transposon Tn4001 carrying aminoglycoside resistance (*aacA-aphD*) gene in *SCCmec* V was noted in 87% (n=264) of the isolates (Table S5). Various cassette chromosome recombinase genes (*ccrAA*, *ccrAB2*, *ccrC*), a high affinity potassium transporter (*kdp*), and a zinc resistance gene (*czrC*) were found to be integrated into *SCCmec* V (5C2 and 5) and *SCCmec* V (5C2) elements (Table S5).

Chromosomal mediated antimicrobial Resistance:

Acquired genes and chromosomal mutations encoding for antimicrobial resistance is shown in Fig 6. Acquisition of Tn552 transposon-mediated beta-lactamase (*blaZ*) conferring penicillin resistance was identified in all of the isolates (Table S5). For the identified copies of *blaZ*, 6% (n=19) were carried on the chromosome, 25% (n=75) were carried on the IRP, and 69% (n=211) were carried in two location (one copy on the chromosome, and another on the IRP). Trimethoprim resistance determinants *dfrG* and *dfrA* were identified in the chromosomes of 99% (n=302) and 100% (n=305) of the isolates respectively. The most common mechanism of fluroquinolone resistance was the double serine mutations; *gyrA* (S84L) and *grlA* (S80Y, S80F). The mutations S84L with S80Y and S84 with S80F accounted for fluroquinolone resistance in 70% (n=215) and 9% (n=28) of the isolates respectively. The double serine mutation (S80F with S84L) was predicted to be replaced by S84L with S80Y in the early 1990s (Fig 4).

Acquired antimicrobial resistance:

All ST772 isolates uniformly carried an IRP, except the basal strains (Fig 6). The IRP contained loci for multiple antibiotic resistant genes: beta-lactam (*blaZ*), aminoglycosides (*aphA3-sat-aadE*), macrolide (*mphC*), macrolide-lincosamide-streptogramin B (*msrA*) and bacitracin (*bacA*, *bacB*) (Fig S3). This genetic mobile element was recognised in 96% (n=292) of the isolates. Aminoglycoside resistant determinants (*aphA3* and *aadE*) were identified in 95% (n=290) of the isolates (Table S5). The genes *mphC* and *msrA* were found in 96% (n=292) and 96% (n=291) of the isolates respectively (Table S5).

Pangenome analysis:

The pangenome analysis of ST772 *S. aureus* genome identified 7428 orthologs with 2040 core genes (27.4%) and 5388 accessory genes (72.6%) (Fig 7). This included 283 soft core genes, 376 shell genes and 4729 cloud genes. Among the accessory genes, 50% (n=2669) were hypothetical proteins. The isolate ERR217355 had high number of accessory genes (n= 2613), compared to others.

A significant proportion of MGEs is derived from ST772 *S. aureus* accessory genome, that includes SCCmec element, prophages, pathogenicity island, plasmids, transposons, and insertion sequence (Fig 8). Overall, in ST772 *S. aureus* genome, MGEs are diverse. Composition of accessory genome analysis revealed that the composition accounts for a total of 2727 CDS (ranging between, 435-2613) in ST772 *S. aureus* genome. Variation in the accessory genome within subclade IIA, particularly results from the loss and gain of SCCmec V elements. A putative pathogenicity island carrying enterotoxin gene *sec* and *sel* was identified in all the genomes.

Discussion:

The epidemiology of MRSA is changing and evolving constantly.⁶³ Several well-characterised MRSA strains are predominant in different regions across the world. The present study revealed the genetic diversity of MRSA dominated by the internationally recognised CCs (CC1, CC22, and CC8). This highlights the risk of MDR and the significant threat of dissemination. In India, PVL positive ST772-SCC*mec* V and ST22-SCC*mec* IV are reported as the first and second most dominant lineages in causing MRSA infections.¹⁶⁻¹⁸ This was confirmed in the present study. These lineages have been reported from community and hospital settings.^{16-18,20,21,64-66} The ST239-SCC*mec* III lineage has been reported as the predominant lineage in many Asian countries.⁶⁷ In the present study, ST239-SCC*mec* III was found as a third most dominant MRSA lineage. Dissemination of ST772 and ST22 MRSA is gradually replacing the ST239-SCC*mec* III lineage in Indian hospitals.⁶⁸

In this study, a high degree of variability was seen among the PVL positive ST22 MRSA affects not only the SCC*mec* marker but also the core genome. The presence of unusual SCC*mec* types in ST22 MRSA indicates an evolutionary genetic change and so demands deeper genomic investigation. Studies have demonstrated the existence of multiple SCC*mec* types in MRSA.⁶⁹⁻⁷¹ In the present study, 6% of the isolates had multiple SCC*mec* types. This suggests that multiple transfer events of SCC*mec* types would resulted in the integration of multiple *ccr* gene complexes in *S. aureus*.^{33,72}

ST772-SCC*mec* V is outcompeting other *S. aureus* lineages in both community and hospital settings in India. Here, we provide evidence for the long term persistence of all subclades (IIA,B,C) of ST772 *S. aureus* and continued causing infections as reported earlier.³⁹ This was evidence by the fact that each of the subclades contained atleast one

isolate collected since early 2000s. Previously, three subclades (ST772-A1, A2, A3) were by defined by Steinig *et al.*³⁹ Of the subclades, the subclade ST772-A1 is reported as old early branching lineage, whilst ST772-A2 and ST772-A3 are defined as dominant and more recent emerging subclades respectively. In this study, Indian ST772 *S. aureus* isolates (including study isolates) were distributed across the clades I and II.

The presence of multiple SCCmec types in the basal strains and isolates of subclade IIA indicates stable core genome, but recombination in SCCmec elements can occur, as reported earlier.⁴¹ This reassures that SCCmec element play a key role in adaptation and rendering the organism fitness to survive. Although, ST772 *S. aureus* has spread internationally, its expansion in India has not been previously studied. Our study provides evidence for the dominance of this lineage in India. In addition, isolates from other countries clustered closely with the Indian isolates, which indicate travel related spread.

The core genome of ST772 *S. aureus* remains stable suggesting the strength of these genomes in adapting to evolutionary pressures for persistence. The substitution rate of ST772 *S. aureus* (1.16×10^{-6}) estimated in this study, is similar to the previous finding which reported 1.18×10^{-6} substitutions/site/year.³⁹ The substitution rate of ST772 *S. aureus* lineage is much lower than the globally dominant *S. aureus* lineages ST22 (1.3×10^{-6}), ST239 (1.6×10^{-6}) and USA 300 clone (1.3×10^{-6}).^{23,73,74} ST772 *S. aureus* lineage has relatively stable core genome and for evidence of microevolution. Our data shows temporal introduction of IRP with the sequential acquisition of SCCmec V (5C2) and accumulation of double serine mutation in 1987 leading to the expansion of multi-drug resistant clones in 1987. However, in the previous study, the ST772 *S. aureus* has been reported to be emerged on the Indian subcontinent in the early 1960s and an

emergence of dominant clade and its subgroups has reported in the early 1990s.³⁹ In the present study, accuracy of MRCA prediction is improved using precise Bayesian model suggested by Bayes factor. Our study indicates that the earliest isolate (2004) was from South Asia (India and Bangladesh), which supports the earlier hypothesis suggesting that South Asia was the likely origin of this lineage.³⁹ Acquisition of IRP, a variant of SCC*mec* V (5C2) and the presence of double serine mutation (*grlA*-S80Y, *gyrA*-S84L) in the dominant clade II isolates were consistent with previous findings and could be the reason behind increased fitness and global expansion of this lineage.³⁹

Studies have emphasized the contribution of mobile genetic elements (MGEs) in the emergence of ST772-SCC*mec* V as a multi-drug resistant and highly virulent lineage.^{20,21,41} The reference genome DAR4145, a representative of ST772 contains SCC*mec* V (5C2), PVL carrying Φ IND772 prophage, pathogenicity island (*vSa*-alpha, beta, and gamma) and an IRP. In the present study, comparative genomic analysis showed that these genetic elements are the conserved molecular markers of ST772 lineage, with few exceptions. Furthermore, comparative analysis of accessory genome showed that MGEs contributed for significant proportion of ST772 *S. aureus* genome. This observation reflects the contribution of MGEs in evolution and dissemination as described in ST22-SCC*mec* IV lineage²³ and USA clones (300, 500, 500 like).⁷⁵

A landmark study on the genomic investigation of the ST772 *S. aureus* detailed the key factors involved in its expansion, evolution and intercontinental transmission. This includes acquisition of IRP, fixation of a smaller SCC*mec* V (5C2) and replacement of mutation in *grlA* (S80F to S80Y).³⁹ A similar findings was confirmed in the present study. An extrachromosomal plasmid in USA300 and an integrated element in the SCC*mec* of ST80-SCC*mec* IV lineage shared a similar genetic backbone of the IRP of

ST772-SCC*mec* V lineage.³⁹ The substitution S84L with S80Y is reported as the most common double serine mutation in *S. aureus*.⁷⁶ Acquisition of PVL carrying Φ IND772 prophage and the pathogenicity islands has occurred even before the emergence of ST772-SCC*mec* V lineage. No significant difference in the cellular toxicity or biofilm formation have been reported between the basal strains and a single globally disseminated clade (ST772-A).³⁹ Multi-drug resistance could have provided a strong fitness advantage for the ST772 lineage in hospital settings.

Multiple intercontinental spread from India to various countries in Europe, Australia and Asia have been reported.³⁹ In addition, outbreaks in the NICU in Ireland and a hospital in Norway have also documented.^{77,78} However, intense endemic dissemination of this clone is not evident following intercontinental spread. ST772-SCC*mec* V has a similar pattern of dissemination as with that of other CA-MRSA such as USA300 clone and the ST80-SCC*mec* IV lineage.

In conclusion, ST772, ST22 and ST239 were observed as the predominant STs. ST772-SCC*mec* V was seen as the well-established lineage. Whilst, the occurrence of ST22 with various SCC*mec* types suggests an evolutionary genetic change and so demands deeper genomic investigation. ST772-SCC*mec* V (Bengal bay clone) was emerged on the Indian subcontinent in 1959 and diverged into a dominant clade in 1987. The acquisition of IRP, SCC*mec* V (5C2) and the double serine mutations associated with fluroquinolone resistance are likely to be the important attributes to the success of the ST772 lineage. The clinical antimicrobial use pattern may have driven the spread and survival of ST772 MRSA in hospitals. ST772-SCC*mec* has the multi-drug resistance trait of HA-MRSA and the epidemiological characteristics of CA-MRSA.

489 **Acknowledgements:** None

490 **Funding:** No specific funding has been received for this study.

491 **Transparency declarations:** None to declare

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

References:

1. Jevons M. Celbenin-resistant staphylococci. *Br. Med. J* 1961; 1: 124–125.
2. Robinson DA, Enright MC. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2003;47(12):3926-34.
3. Boyle-Vavra S, Daum RS. Community-acquired methicillin-resistant *Staphylococcus aureus*: the role of Panton-Valentine leukocidin. *Lab Invest* 2007;87(1):3-9.
4. Donker GA, Deurenberg RH, Driessen C *et al*. The population structure of *Staphylococcus aureus* among general practice patients from The Netherlands. *Clin Microbiol Infect* 2009;15(2):137-43.
5. Huh K, Chung DR. Changing epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in the Asia-Pacific region. *Expert Rev Anti Infect Ther* 2016;14(11):1007-1022.
6. Chua K, Laurent F, Coombs G, *et al*. Antimicrobial resistance: Not community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA)! A clinician's guide to community MRSA - its evolving antimicrobial resistance and implications for therapy. *Clin Infect Dis* 2011;52(1):99-114.
7. Rajkumar S, Sistla S, Manoharan M, *et al*. Prevalence and genetic mechanisms of antimicrobial resistance in *Staphylococcus* species: A multicentre report of the indian council of medical research antimicrobial resistance surveillance network. *Indian J Med Microbiol* 2017;35(1):53-60.

8. Veeraraghavan B, Walia K. Antimicrobial susceptibility profile & resistance mechanisms of Global Antimicrobial Resistance Surveillance System (GLASS) priority pathogens from India. *Indian J Med Res* 2019;149(2):87-96.
9. Nadig S, Namburi P, Raghunath D *et al.* Genotyping of methicillin-resistant *Staphylococcus aureus* isolates from Indian hospitals. *Curr Sci* 2006;91:1364–9.
10. Arakere G, Nadig S, Ito T *et al.* A novel type-III staphylococcal cassette chromosome *mec* (SCCmec) variant among Indian isolates of methicillin-resistant *Staphylococcus aureus*. *FEMS Microbiol Lett* 2009;292:141–8.
11. Gadepalli R, Dhawan B, Kapil A *et al.* Clinical and molecular characteristics of nosocomial methicillin-resistant *Staphylococcus aureus* skin and soft tissue isolates from three Indian hospitals. *J Hosp Infect* 2009;73:253–63.
12. Goering RV, Shawar RM, Scangarella NE *et al.* Molecular epidemiology of methicillin-resistant and methicillin susceptible *Staphylococcus aureus* isolates from global clinical trials. *J Clin Microbiol* 2008;46:842–7
13. Afroz S, Kobayashi N, Nagashima S *et al.*, Genetic characterization of *Staphylococcus aureus* isolates carrying Panton-Valentine leukocidin genes in Bangladesh. *Jpn J Infect Dis* 2008;61(5):393-6.
14. Kulkarni GB, Pal PK, Veena Kumari HB *et al.* Community-acquired methicillin-resistant *Staphylococcus aureus* pyomyositis with myelitis: a rare occurrence with diverse presentation. *Neurol India* 2009;57:653–6.
15. Bouchiat C, El-Zeenni N, Chakrakodi B *et al.* Epidemiology of *Staphylococcus aureus* in Bangalore, India: emergence of the ST217 clone and high rate of resistance to erythromycin and ciprofloxacin in the community. *New Microbes New Infect* 2015;7:15–20.

16. Dhawan B, Rao C, Udo EE *et al.* Dissemination of methicillin-resistant *Staphylococcus aureus* SCCmec type IV and SCCmec type V epidemic clones in a tertiary hospital: challenge to infection control. *Epidemiol Infect* 2015;143:343–53.
17. D'souza N, Rodrigues C, Mehta A. Molecular characterization of methicillin-resistant *Staphylococcus aureus* with emergence of epidemic clones of sequence type (ST) 22 and ST 772 in Mumbai, India. *J Clin Microbiol* 2010;48:1806–11.
18. Shambat S, Nadig S, Prabhakara S *et al.* Clonal complexes and virulence factors of *Staphylococcus aureus* from several cities in India. *BMC Microbiol* 2012;12:64.
19. Bakthavatchalam YD, Babu P, Munusamy E *et al.* Genomic insights on heterogeneous resistance to vancomycin and teicoplanin in Methicillin-resistant *Staphylococcus aureus*: A first report from South India. *PLoS One* 2019; 14(12):e0227009.
20. Monecke S, Baier V, Coombs GW *et al.* Genome sequencing and molecular characterisation of *Staphylococcus aureus* ST772-MRSA-V, "Bengal Bay Clone". *BMC Res Notes* 2013;6:548.
21. Steinig EJ, Andersson P, Harris SR *et al.* Single-molecule sequencing reveals the molecular basis of multidrug-resistance in ST772 methicillin-resistant *Staphylococcus aureus*. *BMC Genomics* 2015;16:388.
22. Harris SR, Feil EJ, Holden MT *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010;327(5964):469-74.
23. Holden MT, Hsu LY, Kurt K *et al.* A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* 2013;23(4):653-64.

24. Strauß L, Stegger M, Akpaka PE et al. Origin, evolution, and global transmission of community-acquired *Staphylococcus aureus* ST8. *Proc Natl Acad Sci U S A* 2017;114(49):E10596-E10604.
25. Harrison EM, Coll F, Toleman MS et al. Genomic surveillance reveals low prevalence of livestock-associated methicillin-resistant *Staphylococcus aureus* in the East of England. *Sci Rep* 2017;7(1):7406.
26. Winn WC. Koneman's color atlas and textbook of diagnostic microbiology. Lippincott Williams & Wilkins, 2006.
27. Zhang K, McClure JA, Elsayed S *et al.* Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 2005 ;43(10):5026-33.
28. Ghaznavi-Rad E, Nor Shamsudin M, Sekawi Z *et al.* A simplified multiplex PCR assay for fast and easy discrimination of globally distributed staphylococcal cassette chromosome *mec* types in methicillin-resistant *Staphylococcus aureus*. *J Med Microbiol* 2010;59(Pt 10):1135-9.
29. Enright MC, Day NP, Davies CE *et al.* Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 2000;38(3):1008-15.
30. Harmsen D, Claus H, Witte W *et al.* Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. *J Clin Microbiol* 2003;41(12):5442-8.
31. Aziz RK, Bartels D, Best AA *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008; 9(1):75

32. Wattam AR, Abraham D, Dalay O *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 2013; 42(D1):D581–D591
33. Kaya H, Hasman H, Larsen J *et al.* SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. *mSphere* 2018;3(1). pii: e00612-17.
34. Chen L, Yang J, Yu J *et al.* VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005; 33(suppl_1):D325–D328.
35. Jia B, Raphenya AR, Alcock B *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database, *Nucleic Acids Res* 2016; 45(D1):D566–D573.
36. Bradley P, Gordon NC, Walker TM *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* 2015;6:10063.
37. Arndt D, Grant JR, Marcu A *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016; 44(W1):W16–W21.
38. Bertelli C, Laird MR, Williams KP *et al.* IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017; 45(W1): W30-35.
39. Steinig EJ, Duchene S, Robinson DA *et al.* Evolution and Global Transmission of a Multidrug-Resistant, Community-Associated Methicillin-Resistant Staphylococcus aureus Lineage from the Indian Subcontinent. *mBio* 2019;10(6). pii: e01105-19.
40. Prabhakara S, Khedkar S, Shambat SM, Srinivasan R, Basu A, Norrby-Teglund A, *et al.* Genome sequencing unveils a novel sea enterotoxin-carrying PVL phage in Staphylococcus aureus ST772 from India. *PLoS One*. 2013;8, e60013.

41. Balakuntla J, Prabhakara S, Arakere G. Novel rearrangements in the
staphylococcal cassette chromosome mec type V elements of Indian ST772 and
ST672 methicillin resistant *Staphylococcus aureus* strains. PLoS One. 2014;9,
e94293.
42. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm
and its applications to single-cell sequencing. J comput Biol. 2012;19(5):455-77
43. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein
function. Nucleic Acids Res 2003;31(13):3812-4.
44. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler
transform. Bioinformatics 2009;25(14):1754-60.
45. Seemann T. Snippy: fast bacterial variant calling from NGS
reads. <https://github.com/tseemann/snippy>. Accessed 12 March 2020
46. Garrison E, Marth G. Haplotype-based variant detection from short-read
sequencing. arXiv 2012. arXiv:1207.3907.
47. Croucher NJ, Page AJ, Connor TR *et al.* Rapid phylogenetic analysis of large
samples of recombinant bacterial whole genome sequences using Gubbins.
Nucleic acids research 2015;43(3):e15-.
48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
analysis of large phylogenies. Bioinformatics 2014;30(9):1312-3.
49. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and
annotation of phylogenetic and other trees. Nucleic acids research.
2016;44(W1):W242-5.
50. Cheng L, Connor TR, Sirén J *et al.* Hierarchical and spatially explicit clustering of
DNA sequences with BAPS software. Molecular biology and evolution
2013;30(5):1224-8.

51. Tonkin-Hill G, Lees JA, Bentley SD *et al.* RhierBAPS: an R implementation of the population clustering algorithm hierBAPS. Wellcome open research 2018; 3:93.
52. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 2007;7(1):214.
53. Rambaut A, Drummond AJ. LogCombiner v1. 8.2. LogCombinerv1.8.2. 2015, <http://beast.bio.ed.ac.uk>.
54. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology. 2018 Sep;67(5):901.
55. Li WL, Drummond AJ. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Molecular biology and evolution. 2012 Feb 1;29(2):751-61.
56. Helfrich P, Rieb E, Abrami G, Lücking A, Mehler A. TreeAnnotator: versatile visual annotation of hierarchical text relations. InProceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) 2018 May.
57. Rambaut A. FigTree v1.4.2, A Graphical Viewer of Phylogenetic Trees. 2014, Available from <http://tree.bio.ed.ac.uk/software/figtree/>
58. Rambaut A, Drummond AJ, Xie D *et al.* Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology 2018;67(5):901.
59. Page AJ, Cummins CA, Hunt M *et al.* Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31(22):3691-3.
60. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014 ;30(14):2068-9.
61. Ozer EA, Allen JP, and Hauser AR. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGent. BMC Genomics 2014 15:737

62. HUERTA-CEPAS, Jaime, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular biology and evolution*, 2017, 34.8: 2115-2122.
63. Monecke S, Coombs G, Shore AC *et al.* A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. *PLoS One* 2011;6(4):e17936.
64. Pokhrel RH, Aung MS, Thapa B *et al.* Detection of ST772 Panton-Valentine leukocidin-positive methicillin-resistant *Staphylococcus aureus* (Bengal Bay clone) and ST22 *S. aureus* isolates with a genetic variant of elastin binding protein in Nepal. *New Microbes New Infect* 2016;11:20-7.
65. Pinto AN, Seth R, Zhou F *et al.* Emergence and control of an outbreak of infections due to Panton-Valentine leukocidin positive, ST22 methicillin-resistant *Staphylococcus aureus* in a neonatal intensive care unit. *Clin Microbiol Infect* 2013;19(7):620-7.
66. Goudarzi M, Navidinia M, Beiranvand E *et al.* Phenotypic and Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* Clones Carrying the Panton-Valentine Leukocidin Genes Disseminating in Iranian Hospitals. *Microb Drug Resist*. 2018; doi: 10.1089/mdr.2018.0033. [Epub ahead of print]
67. Monecke S, Slickers P, Gawlik D *et al.* Molecular Typing of ST239-MRSA-III From Diverse Geographic Locations and the Evolution of the SCCmec III Element During Its Intercontinental Spread. *Front Microbiol* 2018;9:1436.
68. Sunagar R, Hegde NR, Archana GJ *et al.* Prevalence and genotype distribution of methicillin-resistant *Staphylococcus aureus* (MRSA) in India. *J Glob Antimicrob Resist* 2016;7:46-52.

69. Bhutia KO, Singh TSK, Adhikari L *et al.* Molecular characterization of community- & hospital-acquired methicillin-resistant & methicillin-sensitive *Staphylococcus aureus* isolates in Sikkim. *Indian J Med Res* 2015;142:330–335
70. Murugesan S, Perumal N, Mahalingam SP *et al.* Analysis of antibiotic resistance genes and its associated SCCmec types among nasal carriage of methicillin resistant coagulase negative *Staphylococci* from community settings, Chennai, Southern India. *J Clin Diagnostic Res* 2015;9:DC01–DC05
71. Nagasundaram N, Sistla S. Existence of multiple SCCmec elements in clinical isolates of methicillin-resistant *Staphylococcus aureus*. *J Med Microbiol* 2019;68(5):720-727.
72. Jamrozy DM, Harris SR, Mohamed N *et al.* Pan-genomic perspective on the evolution of the *Staphylococcus aureus* USA300 epidemic. *Microb Genom* 2016 May 31;2(5):e000058.
73. Baines SL, Holt KE, Schultz MB, Seemann T, Howden BO, Jensen SO *et al.* Convergent adaptation in the dominant global hospital clone ST239 of methicillin-resistant *Staphylococcus aureus*. *mBio*. 2015 Mar 3;6(2):e00080.
74. Glaser P, Martins-Simões P, Villain A, Barbier M, Tristan A, Bouchier C *et al.* Demography and Intercontinental Spread of the USA300 Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Lineage. *mBio*. 2016;7(1):e02183-15.
75. Jamrozy DM, Harris SR, Mohamed N, Peacock SJ, Tan CY, Parkhill J, Anderson AS, Holden MTG. Pan-genomic perspective on the evolution of the *Staphylococcus aureus* USA300 epidemic. *Microb Genom*. 2016 May 31;2(5):e000058
76. Yamuna DB, Abirami S, Rajeshwari M *et al.* Levonadifloxacin, a recently approved benzoquinolizine fluoroquinolone, exhibits potent in vitro activity against

contemporary *Staphylococcus aureus* isolates and Bengal Bay clone isolates AQ1 collected from a large Indian tertiary care hospital. *J Antimicrob Chemother* 2020; doi:10.1093/jac/dkaa142 (ahead of print)

77. Brennan GI, Shore AC, Corcoran S *et al.* Emergence of hospital- and community-associated panton-valentine leukocidin-positive methicillin-resistant *Staphylococcus aureus* genotype ST772-MRSA-V in Ireland and detailed investigation of an ST772-MRSA-V cluster in a neonatal intensive care unit. *J Clin Microbiol* 2012;50(3):841-7.

78. Blomfeldt A, Larssen KW, Moghen A *et al.* Bengal Bay clone ST772-MRSA-V outbreak: conserved clone causes investigation challenges. *J Hosp Infect* 2017 Mar;95(3):253-258.

Fig 1: a) Maximum likelihood phylogeny was constructed based on the seven house-keeping genes of *S. aureus*. The tree was marked with the meta-data using interactive tree of life (iTOL.v.4). The inner most ring represents clonal complexes (CCs), the second ring denotes sequence types (STs), the third ring represents SCCmec types and the outer most ring represents the *spa* types. CCs, STs and SCCmec types are colour coded as mentioned in legends.

Fig 2: a) Maximum likelihood phylogenetic tree was constructed based on the core genome SNP of ST772 *S. aureus* lineage (n=305). The innermost ring represents BAPS clusters. The second, third, fourth and fifth rings denotes the presence/absence of *mecA* gene, presence of various SCCmec types, presence/absence of integrated resistance plasmid (IRP) and the PVL gene respectively. Isolates in the clades are colour coded; study isolates are in blue color and other Indian isolates are in red color. b) Bayesian Analysis of Population Structure (BAPS) analysis of ST772 *S. aureus* isolates showed the presence of two distinct sequence clusters (SC) based on the core SNPs. The basal strains and isolates of subclade IIC (yellow) are classified in SC1, while isolates of subclade IIb (green) are represented in SC2.

Fig 3: Root to tip genetic distance of the ST772 *S. aureus* isolates. The total number of SNPs present within each isolate compared to that in the index isolate was computed and plotted against the date on which each isolate was recovered. The graph includes regression lines with 95% CI for each group. Sizes of circles correspond to the number of isolates.

Fig 4: Evolutionary analysis of ST772 *S. aureus*. The maximum clade credibility tree was generated from the filtered single-nucleotide polymorphisms (SNPs) in the core genome of 305 *S. aureus* ST772 genome using BEAST analysis.

Fig 5: The Baesiyan skyline plot (BSP) analysis of ST772 *S. aureus* effective population size and it's time for the most recent common ancestor. BSP was based on the strict clock' coalescent framework analysis and was constructed using 305 sequences (representing all years and countries). X-axis represents time in years and Y-axis shows the effective population size. The blue thick line represents the median, while the blue band represents 95% highest posterior density (HPD) intervals.

Fig 6: Distribution of antimicrobial resistant (AMR) genes and virulence genes identified among the clades of ST772 *S. aureus* isolates. The presence of acquired AMR genes as well as chromosomal mutations is denoted in red color. The presence of virulence genes are indicated in blue colour.

Fig 7: Pangenome analysis of ST772 *S. aureus* isolates using roary pipeline and visualised using Phandango, an interactive viewer for bacterial population genomics. Blue bar indicates the pan-genome of ST772 *S. aureus* isolates with 7428 annotated genes detected in all the analysed genomes. The maximum likelihood phylogenetic tree was integrated alongside with the presence and absence of genes. Each row corresponds to an isolate and each column corresponds to the variation in genes. The presence of genes is showed in the decreasing frequency. Empty space indicates the absence of genes.

Fig 8: Mobile genetic elements identified in the accessory genome of ST772 *S. aureus* isolates.

Fig S1: Maximum likelihood phylogenetic tree was constructed based on the core genome SNP of ST772 *S. aureus* lineage (n=304). The year and country of isolation of ST772 *S. aureus* isolates.

794 **Fig S2:** Pathogenicity island identified in ST772-SCCmec V lineage. a) vSa-alpha
 795 pathogenicity island carried various exotoxins and superantigens. b) vSa-beta
 796 pathogenicity island harboured enterotoxin gene cluster. c) vSa-gamma genomic island
 797 comprised of alpha hemolysin and exfoliative (*eta*) toxin.

798 **Fig S3:** Genetic environment of integrated resistance plasmid (IRP) in VB9352
 799 (accession no. CP035670)

800

801

Clonal Complex

- CC1
- CC121
- CC22
- CC30
- CC5
- CC580
- CC672
- CC72
- CC8
- CC88
- Singleton

Sequence types

- ST1
- ST772
- ST1358
- ST649
- ST1037
- ST217
- ST22
- ST2371
- ST3976
- ST636
- ST896
- ST1482
- ST382
- ST1290
- ST1307
- ST6
- ST801
- ST580
- ST2775
- ST361
- ST672
- ST72
- ST1803
- ST239
- ST3324
- ST368
- ST630
- ST8
- ST88
- ST616
- ST1947
- ST1598

SCCMec types

- I
- II
- III
- IVa
- IVc
- IVd
- V
- III/V

a)

Tree scale: 0.0001

BAPS

- 1
- 2

mec A

- Positive
- Negative

SCCmec types

- mecA
- PseudoSCCmec
- SCCmec V (5C2) combinations
- SCCmec (5C2 & 5) combinations
- SCCmec IV remnant
- SCCmec (5C2 & 5)
- SCCmec V (5C2)
- SCCmec negative

IRP

- Positive
- Negative

PVL

- Positive
- Negative

b)

- Double mutation (*grlA*_S80Y, *gyrA*_S84L)
- Double mutation (*grlA*_S80F, *gyrA*_S84L)
- Single mutation (*grlA*_S80F)
- Single mutation (*gyrA*_S84L)

Pathogenicity island:

- vSa-alpha
- vSa-beta
- vSa-gamma

SCC*mec* V (5C2)
1994 (95% HPD, 1992 - 1997)

Integrated resistance plasmid (IRP)

1991 (95% HPD, 1989-1994)

SCC*mec* V (5C2 & 5)/

SCC*mec* V (5C2)

MRCA: 1968 (95% HPD, 1963-1974)

PVL carrying Φ IND772 prophage

SCC*mec* V (5C2 & 5)

Tree scale: 0.0001

Total genes:7428

Core genes:2040 (27.4%)

Accessory genes:5388 (72.6%)

Tree scale: 0.01

Tree scale: 0.0001

Year of isolation

- 2004-2006
- 2007-2009
- 2010-2012
- 2013-2015
- 2016-2018

Country

- Australia
- Bangladesh
- Denmark
- England
- Germany
- Hong Kong
- India
- Ireland
- Italy
- Netherlands
- New Zealand
- Norway
- Scotland
- United Arab Emirates

a) vSa alpha

b) vSa beta

c) vSa gamma

Abortive infection bacteriophage resistance protein

Hypothetical protein

ATP binding cassette transporter

HsdM

HsdS1

Extracellular calcium binding domain

Deoxyribonucleotide triphosphate pyrophosphatase

Putative metallophosphoesterase

ISSau3 transposase

Superantigen like protein

Ornithine carbamoyltransferase (argF)

Carbamate kinase (arcC1)

Ribonucleotide phosphohydrolase

Acetyltransferase (GNAT) family protein