Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

WHOLE GENOME SEQUENCING AND PANGENOMIC STUDY OF MULTIDRUG-RESISTANT *STAPHYLOCOCCUS AUREUS* ISOLATES FROM PESHAWAR, PAKISTAN

HABIB ULLAH KHAN
Department Biotechnology, Abdul Wali Khan University Mardan, Pakistan. Email: janabad555@gmail.com

FAZAL HANAN
Department Pathology Saidu Group of Teaching Hospital/Saidu Medical College Swat Pakistan.
Email: drfhanan@gmail.com

AMJAD ALI
Atta Ur Rahman School of Applied Biosciences (ASAB) National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan. Email: amjad.ali@asab.nust.edu.pk

ABID UL GHAFOOR
District Medical Specialist Cat C Hospital Khwaza Khaila Swat Pakistan.
Email: Dr.abidulghafoor1@gmail.com

AIZAZ ALI
Department Biotechnology, Abdul Wali Khan University Mardan, Pakistan. Email: aliaizazhk@gmail.com

KARIM GUL
Department Biotechnology, Abdul Wali Khan University Mardan, Pakistan.
Email: karimgul@awkum.edu.pk

Dr. FAZAL JALIL *
Associate Professor, Department Biotechnology, Abdul Wali Khan University Mardan, Pakistan.
*Corresponding Author Email: fazaljalil@awkum.edu.pk

Abstract

We performed Whole-genome sequencing (WGS) of six *S. aureus* strains isolated from Peshawar, Pakistan, was conducted in this present study to identify the possible pangenome of *S. aureus*. These strains were then compared with the 200 fully sequenced genomes and 27 draft genomes of *S. aureus* from Pakistan available in the PATRIC database. Bioinformatics studies involving sequence assembly, quality check, Multi-Locus Sequence Typing (MLST), and Cluster of Orthologous Genes (COGs) were performed to examine the genetic variations and the evolutionary patterns of these strains. The de novo assembly of the *S. aureus* genomes indicated the genome size to lie between 1. 6 to 2. 9 Mb with average GC percentage. MLST analysis identified diverse sequence types among SA1, SA2, SA3, SA4, SA5, SA6 isolates belong to ST 45, 30, 772, 1413, 30, 30 and carries no SCCmec, (mecA gene, SCCmec type V, VII, mecA), mecA, (SCCmec type IV, mecA, Iva), (SCCmec type IV, mecA, Iva). Using the pangenome analysis of the 200 *S. aureus* genomes, it found a total of 9, 118 genes while sharing 1, 620 core genes, 307 soft core genes, 1, 161 shell genes, and the remaining 6, 030 of cloud genes. Thus, this analysis reveals the fact that the pangenome of *S. aureus* is open, suggesting its ability to obtain new genes and adapt to the environment. The COG analysis showed the functional classifications of the genes underlying metabolic processes. These outcomes are crucial for clinicians and therapeutic strategies for treating bacterial *S. aureus* infections.

Keywords: *S. aureus,* MRSA, WGS, MLST.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

1. INTRODUCTION

Staphylococcus aureus (S. aureus) is a major pathogen, which accounts for dramatic nosocomial infections. Methicillin-resistant S. MRSA is one of the very organisms of nosocomial infections that pose a real danger to immunocompromised patients (1). The rate of nasal colonization of *S. aureus* to be roughly 26. 0. 01% in the community and 25. 4% in healthy hospital (2).

Additionally, Bazaid et al observed that during the course of 5 years between January 2015 to December 2019 (3) found 12% of *S. aureus* in patients with urinary tract infections. Moreover, *S. aureus* was also found to be involved in a number of infections including osteoarticular infections, endocarditis, deep soft-tissue infections both in the community and hospital settings, and food poisoning (4-7).

Multidrug-resistant staphylococci may persist in the hospital environment and can neutralize drugs and biocides by biofilm formation or cell conversion in atypical forms (8). The identification of genetic factors that contribute to *S. aureus* pathogenicity and antibiotic resistance is essential for the design of treatment strategies and the prevention of the spread.

Whole genome sequencing(WGS) provides comprehensive genetic map of different organisms; allows to discover the genetic factors contributing to phenotypic distinctions, pathogenicity, as well as understanding of resistance factors (9). Due to comparative analysis of multiple *S. aureus* strains with sequenced complete genomes, the investigators become able to study the evolutionary history and the functional capabilities of the bacterium (10). This type of strategy enables the investigation of bacteria's gene plasticity and unveils not only conserved genes that are required for the organsims's sustainability but also the dispensable genes that can be acquired or lost while housing special functions and virulence attributes (11).

The concept of core and accessory genomes allows to analyze the multifaceted relationships between bacterial species and their evolution, as well as the evolution of clinically relevant phenotypes, including antibiotic resistance (12,13). Comparing the genomes of *S. aureus*, scientists can recognize how exactly populations become resistant to antibiotics and introduce targeted therapies to prevent the distribution of multidrug-resistant strains and decrease the mortality rate among affected patients.

Within the frame of this concept, exists the part of the genome that is conserved in all strains of a given species, which is known as the core genome, and the part of the genome, which differs in strain, or in groups of strains, known as the accessory genome This concept is particularly valuable when studying the genomic potential of *S. aureus* (14). The core genome comprises of genes that are necessary for the cell metabolism and the survival of the bacteria while the dispensable genome has those genes that catalyze versatility in various conditions such as antibiotic resistance and virulence factors. (15). Pangenome analysis hence offers a paradigm for studying the genotypic and phenotypic variation and the evolutionary processes within the species.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

This study therefore aimed at analyzing the genome of six *S. aureus* strains through whole genome sequencing and pangenomics. They were aligned with 200 complete genomic sequences and 27 draft genomes of *S. aureus* from Pakistan. Combining these approaches, we will be able to determine potential candidate genes and to understand the processes underlying the genotypic differences in *S. aureus*. Thus, our results have potential biological therapeutic application and much significance for the clinical prevention and treatment of *S. aureus* infection. Moreover, the improvement of clusters of orthologous groups (COGs) aided in extending the knowledge on various genes and adaptation of these strains. The present study has several significant therapeutic implications and shared several practical significances in the clinical management of *S. aureus* infections.

2. MATERIAL AND METHODS

2.1 Bacterial isolate collection and antibiotic susceptibility testing

S. aureus isolates were taken from Khyber Teaching Hospital in Peshawar. The samples were cultured on mannitol salt agar (MSA) plates for 24 hours at 37°C. The isolates were initially identified using some of the important biochemical tests including oxidase, catalase, and coagulase assays) (16). The antibiotic susceptibility for the following antibiotic classes was determined using the disc diffusion technique in accordance with CLSI 2015 guidelines: Aminoglycosides [gentamicin (10µg) and streptomycin (25µg)], macrolides [erythromycin (15µg)], clindamycin (2µg)], oxazolidinones [linezolid (30µg)], cephalosporins [cefixime (5µg) and cefepime (30µg)], carbapenems [meropenem (10µg)] (17).

2.2 DNA Extraction, quantification and whole genome sequencing

As part of the study titled " as a part of the Technical University of Denmark's (DTU) research project "Two Weeks in the World (TWIW)," The Qiagen DNeasyÂ® Blood & Tissue kit (Qiagen, Venlo, Netherlands) was utilized for DNA extraction, and the Qubit dsDNA highsensitivity (HS) assay kit (Carlsbad, CA, USA) was utilized for quantification) (18). Pairedend sequencing was used to process each sequence on an Illumina NextSeq 500 platform.

2.3 Sequence submission

The genomic sequences were uploaded with project Accession number PRJEB56918 to the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home) online archive.

2.4 Specie identification and sequence retrieval

The species were identified by using FIDBAC database (http://fbac.dmicrobe.cn/), and the raw reads FASTQ files of the isolated strains were taken from the ENA database.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

2.5 Assembly and QC check

The initial stage in assembling the genome was to assess the quality of the raw sequencing data using the QUAST software (19). The genome assembler Hybrid*SPA*des was used assemble the raw reads together (https://github.com/kbaseapps/kb_*SPA*des) (20). One bioinformatic technique that has gained a lot of attention and is used extensively is called Hybrid*SPA*des. It is quite effective at assembling DNA sequences, especially when dealing with complex genomes.

Hybrid*SPA*des are used efficiently for producing the de Bruijn graph with graph-based algorithms. To precisely and consistently evaluate genetic information and assemble genomic sequences this methodology help scientist to assemble bacterial genomes. For assembly number of parameters have important role total count of contigs, genome size, GC content, and total coverage. For predicting these parameters and assembled genome evaluation QUAST (Quality Assessment for Genome Assembly). This tool is intended to be developed for accessing the quality of assembled genome. This tool helps the researchers to provide range of statistics like N50, L50 *etc*., and completeness. This led the researchers to quantitively evaluate their assembly data. The default contig size >500 was accounted to run the assembly.

2.6 Genome completeness identification

To confirm the completeness and annotation of the genomes, CheckM V1.0.18 (21) was utilized to run the genomes following assembly and quality check.

2.7 Genome Annotation

The structural annotation by using the Prokka on the sequenced genomes were subjected with default settings. To precisely annotate bacterial genomes Prokka v1.14.5 (22) were used and this application designed too efficiently. Scientists investigating microbial genomes consider it an essential resource because of its exceptional speed, automation and efficiency.

This process has number of output file formats, such as FASTA for protein and Nucleotide and an Annotation file in GFF3 format which would be an interest. Functional annotation were performed *via* using Rapid Annotation using Subsystem Technology (RAST) server (https://rast.nmpdr.org/) (23). By making the annotation of microbial genomes simpler by RAST The method also aids in the identification of genes, functional elements, and metabolic pathways present in DNA sequences. RAST's automated and curated approaches have been intended to boost our understanding of the genomes and biology of microorganisms.

2.8 MLST

The Sequences were then subjected to MultiLocus Sequence Typing (MLST) analysis Four isolates' contig sequences were examined using MLST Software version 2.0.9 (24). MLST helps researchers in describing genetic diversity in microorganisms e.g., scientists

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

and medical professionals interested in learning about the evolution and epidemiology of microbes, this updated version provides enhanced precision and effectiveness.

The allelic profiles or STs of each strain was determined by using the Institute Pasteur technique (MLSTIP), which made use of eight housekeeping genes (*arcC, arcC, aroE, glpF, gmk, pta, tpi, and yqiL*) in its collection. For epidemiological studies PubMLST (25), or Public Multilocus Sequence Typing, is a potent bioinformatics approach that helps scientists evaluating genetic variants in several genes, which facilitates the tracking and understanding of the transmission pattern of infectious diseases.

PubMLST offers publicly accessible, standardized databases for MLST typing and for *SCCmec* and *spa typing* the assembled genomes subjected to Staphopia*SCCmec* (https://github.com/staphopia/staphopia*SCCmec*) and *spaTyper* (https://github.com/HCGBIGTP/*spa*Typer).

2.9 Pangenome Analysis

The sum of all the genomes is known as the pangenome, which is the entire gene repertoire of a given species, namely the core genome and the dispensable genome. The core genome contains genes critical for the growth of the bacteria and are shared by the genomes of a certain species. While Dispensable genome, not shared by every genome and may be in charge of strain specific traits like pathogenicity, stress tolerance, and sensitivity to different stimuli. The pangenome analysis was performed for six genomes of *S. aureus* isolated strains with the dataset including 27 draft genomes from Pakistan, and 200 complete global genome sequences of *S. aureus* which were downloaded from Pathosystems Resource Integration Centre (PATRIC) (26), a comprehensive bioinformatics resource to download genomic data easily (https://www.bvbrc.org/).

PATRIC was identified as the efficient and most useful database to download genomic sequences on the basis of different filters like host, location, draft or complete genome at once that make it an indispensable instrument for microbiological research. The pangenome analysis on six isolates with predominantly global and local genomes was carried out using Roary version 3.13.0 (27) a pangenome pipeline. Roary reads GFF3 format and performs pangenome analysis. Roary creates three output usable formats and had one graphical output to make easy visualizations.

2.10 COG Analysis

COG database is a strong and adaptable that frequently used in biological research (28). Using this database makes it possible to analyze Clusters of Orthologous Groups (COGs) more effectively, which helps researchers identify functional traits that genes in different organisms share. The COG Bioinformatics tool makes a substantial contribution to the giving of useful insights into gene functions and evolutionary links by facilitating comparative genomics. COG analysis was performed using the Bacterial Pangenome Pipeline B-Pan. For COG, phylogenetic, and pangenome analysis, the B-Pan bacterial pangenome pipeline is used. By using COG analysis, core genes and unique genes participating in several pathways were found.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

3. RESULTS

3.1 Sequence Retrieval

A total of 200 complete genome sequences of *S. aureus* were retrieved from PATRIC reported from around the world on 27th of Dec 2023. Also 27 draft genomes of *S. aureus* from Pakistan were retrieved from PATRIC database on the same date. The *S. aureus* included in the study were all reported as MDR and causing infections in human host.

3.2 Pangenome of Multi Drug Resistant *S. aureus*

Roary estimated a total of 9,118 genes in 200 *S. aureus* genomes, out of which 1620 were core genes that were conserved among 99-100% strains and which is 17.7% of total genes. 307 soft core genes were present among 95-99% strains, 1161 were shell or accessory genes that were present among 15-95% of the strains, and 6030 cloud or unique genes were present among 0-15% of the strains (**Figure 3.1**). The number of genes across genomes and their frequency, as shown in (**Figure 3.2**). The tree is compared among *S. aureus* genomes and is compared to a matrix with the core genes presence and absence (**Figure 3.3**). According to pangenome analysis, when the number of genomes increases, so do the total and unique gene counts (**Figure 3.4 & 3.5**). An open pangenome revealed that the number of *S. aureus* genomes increased along with the overall number of genes.

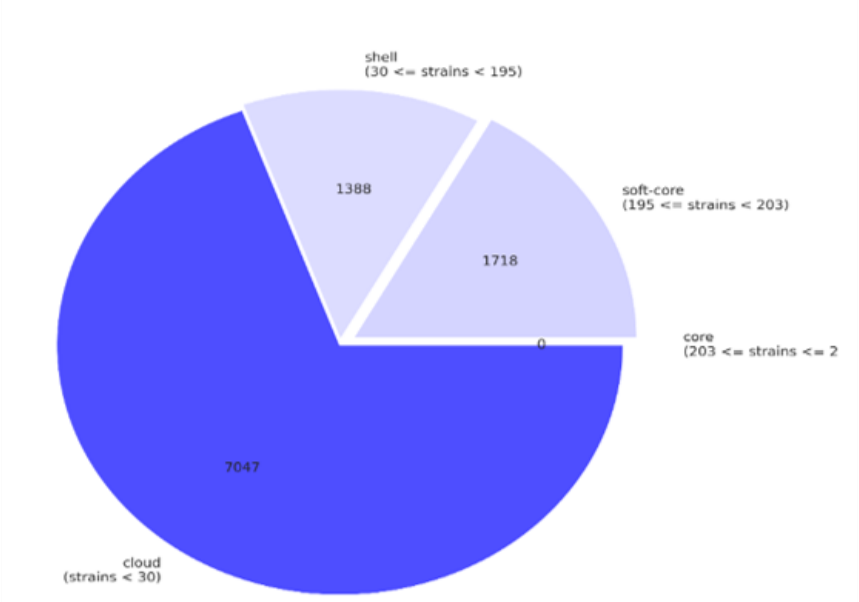

Figure 3.1: *S. aureus* pangenome's global strain pie chart of gene clusters displays Genes classified as core, soft core, and accessory

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

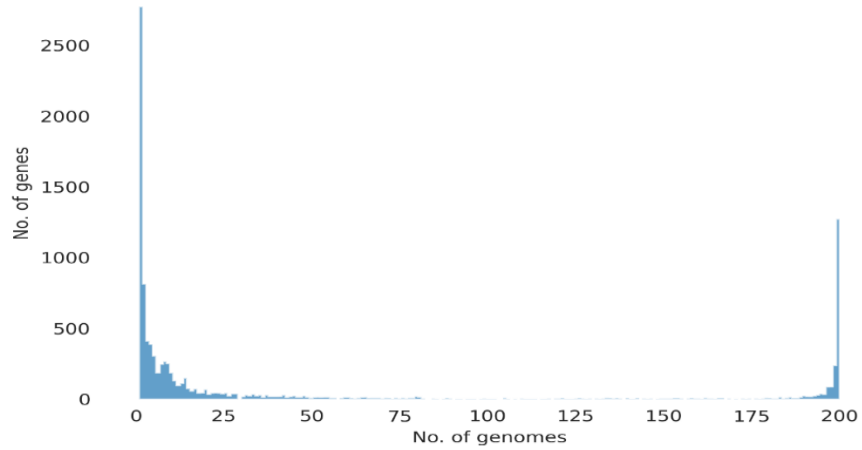

Figure 3.2: *S. aureus* six isolates globally based frequency graph of genome

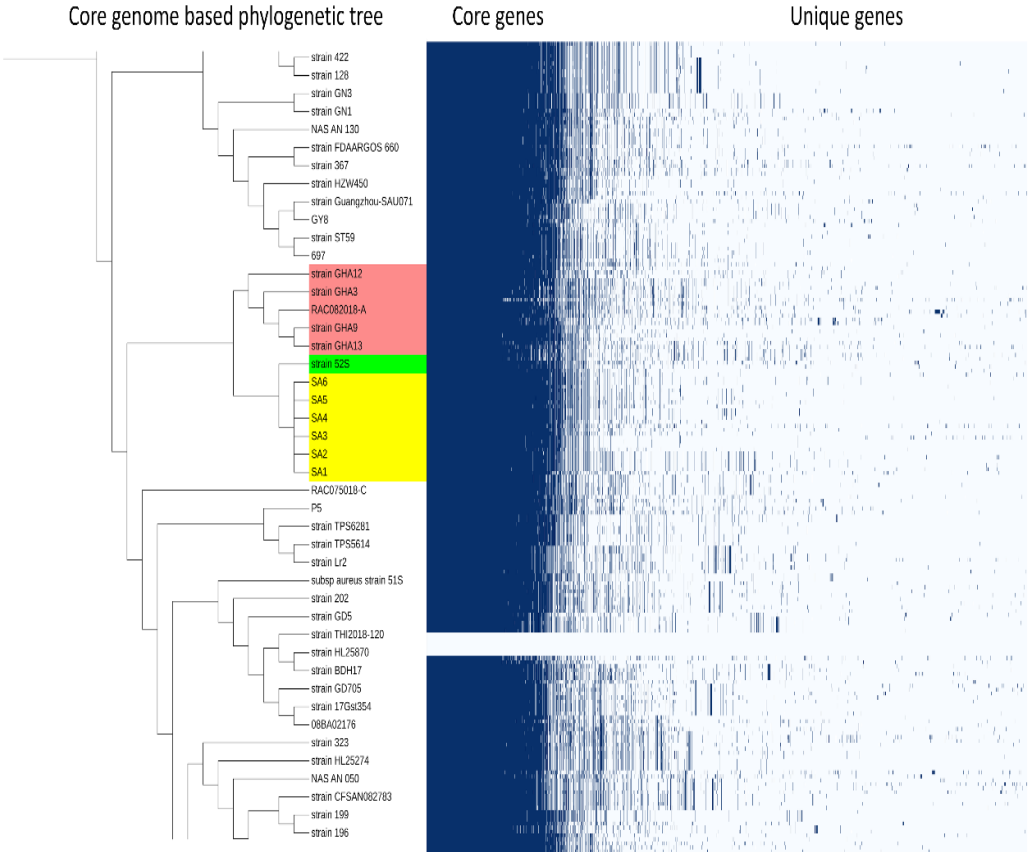

Figure 3.3: Pangenome tree of *S. aureus* in comparison to the matrix showing the presence or absence of core and accessory genes

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

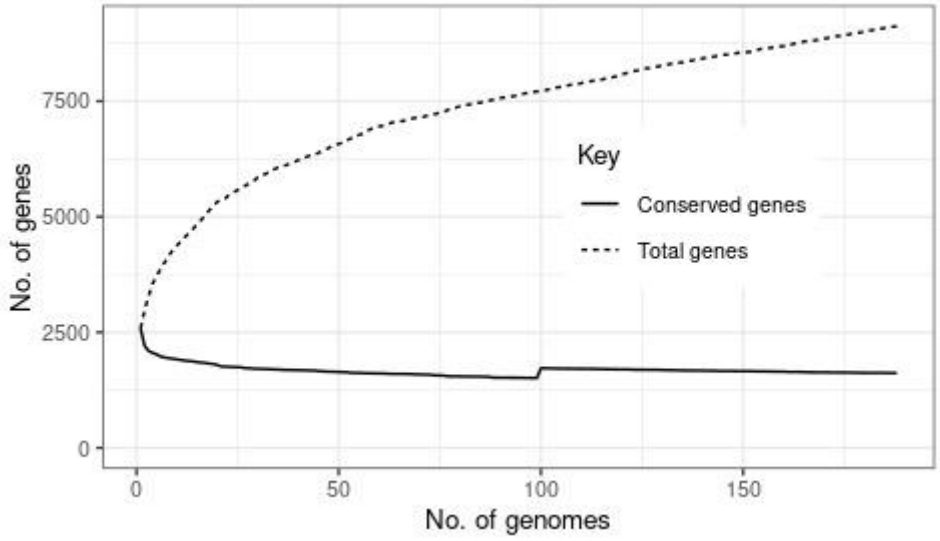

Figure 3.4: A graph illustrating how the overall number of genes increases as the number of *S. aureus* genomes increases

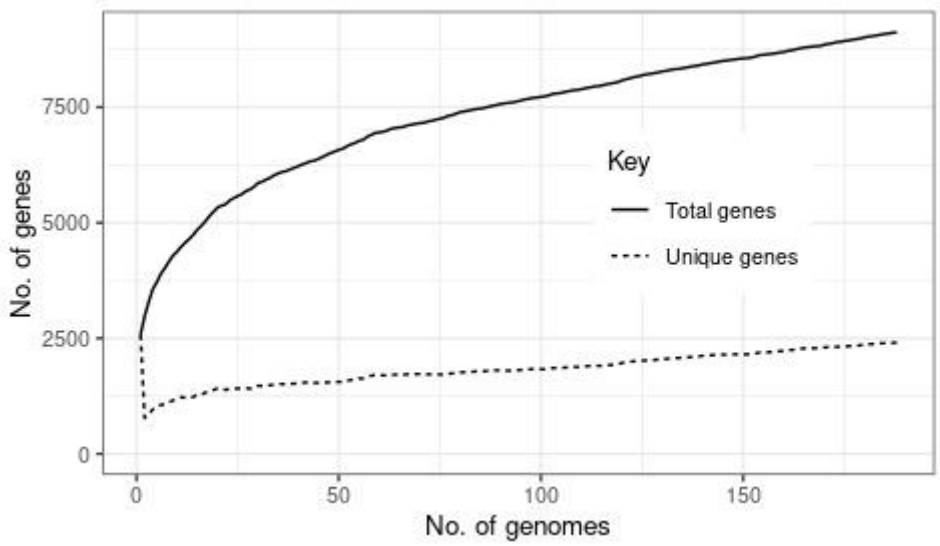

Figure 3.5: Graph shows increase in number of unique genes with increase in number of *S. aureus* Genomes which indicates diversity and adoptability of isolates

3.3 Genomic characteristics

The de novo assembly of SA1, SA2, SA3, SA4, SA5 and SA6 Illumina reads generated 18,68,47,58,38 and correspondingly, 41 contigs (>500 bp) with a GC content ranging from 32.62 to 32.76. The N50 values of SA1, SA2, SA3, SA4, SA5 and SA6 genome are <45k and L50 values are 2 to 11, respectively, and the longest contigs are between the 200k to 1000k bp in length. The genome size of SA1, SA2, SA3, SA4, SA5 and SA6 is between

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

1600k to 2900k bp. The number of predicted CDS are 2570, 2635, 2624, 2682, and 2547. The SA1, SA2, SA3, SA4, SA5 and SA6 harbors are 16 to 32 tRNA genes and 2 to 5 rRNA genes (**Table 3.1**).

Table 3.1: Genomic features and characteristics of MRSA strains SA1, SA2, SA3, SA4, SA5 and SA6

Genomic Characteristics						
	SA1	**SA2**	**SA3**	**SA4**	**SA5**	**SA6**
Genome Size(bp)	2481327	2828872	2819803	2858722	2760883	1604595
Contigs	18	68	47	58	38	41
GC content	32.76	32.72	32.68	32.62	32.69	32.69
N50	415654	82508	15303	134681	170486	136639
L50	2	11	6	7	6	6
CDs	2570	2635	2624	2682	2547	2547
Longest contig size	998731	253525	412914	292957	323688	333155
ST	45	30	772	1413	30	30
tRNA	16	27	28	27	32	24
rRNA	4	4	3	2	4	5
tmRNA	1	1	1	1	1	1

3.4 Multi Locus Sequence Typing of *S. aureus* Isolates

The Pasteur scheme, which consists of the seven housekeeping genes *arcC, aroE, glpF, gmK, pta, tpi, and yqiL*, was used to perform the MLST analysis of all 200 *S. aureus* isolates (**Table 3.2**). MLST analysis revealed 63 different sequence types among 200 strains that show high genetic diversity between *S. aureus* genomes. The most frequent sequence type was ST2 (n=240) but none of the isolated strains belong to sequence type 2. Among other STs shared by *S. aureus* isolates, the most frequently encountered were ST5 (37 isolates), ST8 (33 isolates), ST72 (12 isolates), ST9 (11 isolates), ST30 (10 isolates), and 97 STs has less than 10 isolates (**Table 3.3**). The SA1, SA2, SA3, SA4, SA5, SA6 isolate belong to ST 45, 30, 772, 1413, 30, 30 and carries no *SCCmec*, (mecA gene, *SCCmec* type V, VII, mecA), mecA, (*SCCmec* type IV, mecA, Iva), (*SCCmec* type IV, mecA, Iva). While belongs to variable *spa* type listed in (**Table 3.4**).

Table 3.2: *S. aureus* isolates carry sequence types and allelic profile

Strain Name	MLST (NSTs)	Multi- Locus Allelic Profile							Genome Accession
		arcC	aroE	glpf	gmk	pta	tpi	yqiL	
SA1	45	10	14	8	6	10	3	2	ERR10431539
SA2	30	2	2	2	2	6	3	2	ERR10431540
SA3	772	1	1	1	1	22	1	1	ERR10431542
SA4	1413	6	5	6	2	162	14	5	ERR10431545
SA5	30	2	2	2	2	6	3	2	ERR10431550
SA6	30	2	2	2	2	6	3	2	ERR10431551

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

Table 3.3 Multi Locus Sequence Type of selected *S. aureus*

House- Keeping Genes	Sequence types (STs)	No. of Isolates
	ST5	37
arcC genes	ST8	33
aroE	ST72	12
glpF	ST9	11
gmK	ST30	10
pta	ST45, ST15	8
tpi	ST59	6
yqiL	ST1, ST398, ST22, ST239	5
	ST228, ST152, S772, ST25	4
	ST121, ST7, ST1232, ST105, ST51, ST580	3
	ST188, ST1708, ST630, ST30, ST50	2 each
	ST6, ST88, ST12, ST9, ST20, ST87,	1 each
	ST426, ST59, ST121, ST254, ST113	

Table 3.4: Spa type of *S. aureus* isolates

Sequence name	*spa* Type
SA1	08-16-02-xx-34-13-17-34-16-34
SA2	07-02-16-02-16-17
SA3	t657
SA4	t314
SA5	t363
SA6	t363

3.5 COG Analysis

The COGs gene family has been discovered to be involved in different functions such as amino acid transport, small molecule transport, and morphology. For instance, with strain SA1, SA2, SA3, SA4, SA5, and SA6 (24.7%) Cluster of Orthologous Genes (COGs) for translation, (12.7%) in amino acid metabolism and transport; (10.7%) in coenzyme metabolism; (2.7%) in transcription, and (5.6) are found to be associated with energy production and conversion. There are same pattern goes for all strains but there is a slight difference when it comes to the involvement of COGs in coenzyme metabolism is higher than that of the transcription. Graphical representation of the COGs involvement in various metabolic pathways (**Figure 3.6 & 3.7**)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

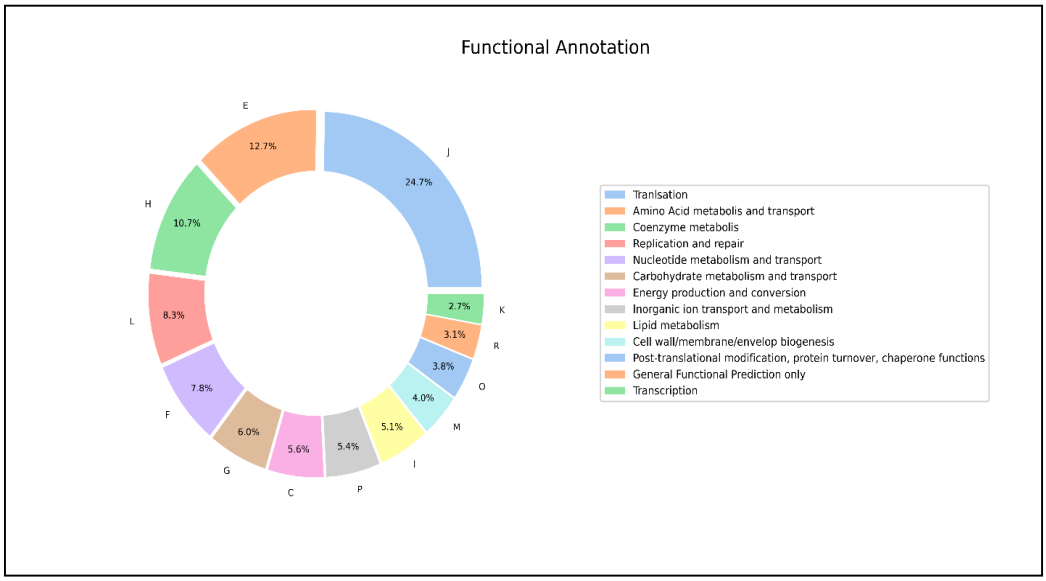

Figure 3.6: Circular visualization of Cluster of orthologous genes in S. aureus of global strains

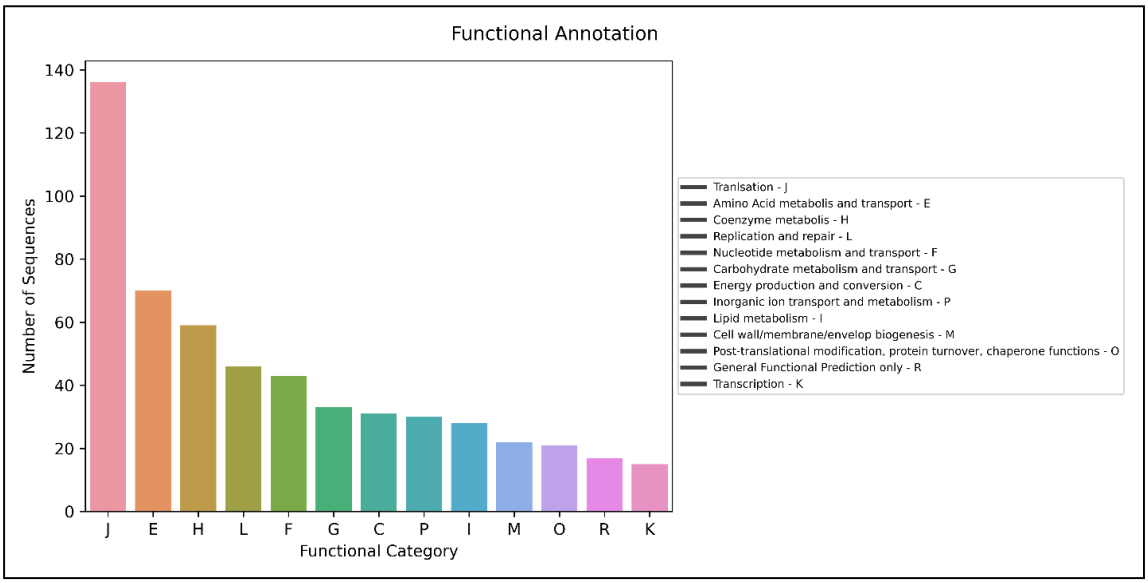

Figure 3.7: Graph shows distribution of Cluster of orthologous genes in *S. aureus* of global strains

4. DISCUSSION

Heavy MRSA is a highly economically burdensome opportunistic multi-resistant disease that emerged due to antibiotic pressure or through MGEs such as plasmids, transposons, and prophages. Its versatile virulence traits include the ability to invade, adhere, and colonize to elude the host's immune system. Furthermore, the management of MRSA

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

infections is made more difficult by the emergence of novel MRSA clones (29). The virulence pattern, antibiotic resistance, and genetic epidemiology of MRSA are crucial for both prevention and treatment. Pakistani genomic data only partially document the molecular epidemiology of MRSA. It is difficult to compare the genomes of Due to the lack of research, MRSA strains in Pakistan are comparable to those in other parts of the world (30-33) have used PFGE, *SCCmec*, and MLST to characterize isolates. No study has performed whole-genome sequencing. From our vast genomic research, we have acquired a detailed understanding of the pathogenicity, resistance mechanisms, and evolutionary relationships of six strains of *S. aureus*. The identification of virulence factors, antibiotic resistance genes, prophage sequences, genomic islands, and whole-genome phylogeny paints a complete picture of these clinically significant isolates.

The WGS of six *S. aureus* isolates from Peshawar exhibited ST 45, 30, 772, 1413, 30 30 respectively. These isolates were phenotypically resistant to Tobramycin, Streptomycin, Amoxillin, Ampicillin, Piparacillin, Penicillin, and Linomycin. These sequence types were previously reported from Rawalpindi, Pakistan and characterization based on *SCCmec* and MLST (Syed et al. 2021).

The in silico *SCCmec* typing revealed that isolates SA2 to SA6 exhibited V, VII, IV and IVa type and all carried mecA gene. Therefore, possibly CA-MRSA (Baig et al. 2018; Lim et al. 2012; Wu et al. 2015) PVL can be virulence-associated gene(s) of the most CA-MRSA strains, that are generally not dependent on β-lactams and are susceptible to non-β-lactams (34-36). While on the contrary HAMRSA are taken up with the hospital and well, in the resistance of non-β-lactam antibiotics and also do not have the PVL gene (37).

In the pangenome analysis conducted with global strains set of 200 *S. aureus* genomes, 1620 were identified as core genes, accounting for 17.7% of total genes, while 307 soft core genes, 1161 were shell or accessory genes found in 15-95% of the strains, and 6030 cloud or unique genes were found in 0-15% of the strains. This depiction of *S. aureus*'s open pangenome displays the organism's ability to constantly acquire genes, emphasizing its genomic flexibility. The link between genome numbers and the development of unique genes highlights the flexibility of organisms (38).

5. CONCLUSION

This work is a great source of genetic opportunities for clinical isolates of *S. aureus* from Peshawer, which can be used to provide epidemiological information and genomic data, also as reference genomes from Pakistan. The types VII, IV and IVa in SA2 - SA6 strains are *SCCmec*, while SA1 does not possess this. The class as well as subtypes identified ST 45, 30, 772, 1413, 30 30 were responsible for the diversity among the isolates. The *S. aureus* isolates show open pangenome that reveal the continuous expansion of the *S. aureus* genome. To have a better understanding of genetic evolution and transmission patterns, more study is required on a sizable MRSA collection in Pakistan. This will improve monitoring efforts and provide useful data on public health.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

References

1) Samia, N.I., Robicsek, A., Heesterbeek, H. and Peterson, L.R. (2022) Methicillin-resistant staphylococcus aureus nosocomial infection has a distinct epidemiological position and acts as a marker for overall hospital-acquired infection trends. *Scientific reports*, **12**, 17007.

2) Alghaithy, A., Bilal, N., Gedebou, M. and Weily, A. (2000) Nasal carriage and antibiotic resistance of Staphylococcus aureus isolates from hospital and non-hospital personnel in Abha, Saudi Arabia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **94**, 504-507.

3) Bazaid, A.S., Saeed, A., Alrashidi, A., Alrashidi, A., Alshaghdali, K., A Hammam, S., Alreshidi, T., Alshammary, M., Alarfaj, A. and Thallab, R. (2021) Antimicrobial surveillance for bacterial uropathogens in Ha'il, Saudi Arabia: A Five-year multicenter retrospective study. *Infection and Drug Resistance*, 1455-1465.

4) Ash, S. and Kennedy, L.E. (2022) *Deep soft-tissue infections*. Oxford University Press Oxford, UK.

5) Bourget, M., Pasquie, M., Charbonneau, H. and Bonnet, E. (2022) Comparable clinical course between coagulase-negative staphylococcal and Staphylococcus aureus endocarditis. *Infection*, **50**, 483-490.

6) Moutaouakkil, K., Abdellaoui, H., Arhoune, B., Atarraf, K., El Fakir, S., Yahyaoui, G., Mahmoud, M., Afifi, M.A. and Oumokhtar, B. (2022) Paediatric osteoarticular infections caused by Staphylococcus aureus producing panton–valentine leucocidin in morocco: Risk factors and clinical features. *African Journal of Paediatric Surgery*, **19**, 78-82.

7) Zhang, F., Wu, S., Lei, T., Wu, Q., Zhang, J., Huang, J., Dai, J., Chen, M., Ding, Y. and Wang, J. (2022) Presence and characterization of methicillin-resistant Staphylococcus aureus co-carrying the multidrug resistance genes cfr and lsa (E) in retail food in China. *International Journal of Food Microbiology*, **363**, 109512.

8) Guo, H., Tong, Y., Cheng, J., Abbas, Z., Li, Z., Wang, J., Zhou, Y., Si, D. and Zhang, R. (2022) Biofilm and small colony variants—an update on staphylococcus aureus strategies toward drug resistance. *International journal of molecular sciences*, **23**, 1241.

9) Montelongo, C., Mores, C.R., Putonti, C., Wolfe, A.J. and Abouelfetouh, A. (2022) Whole-genome sequencing of Staphylococcus aureus and Staphylococcus haemolyticus clinical isolates from Egypt. *Microbiology Spectrum*, **10**, e02413-02421.

10) Young, B.C., Golubchik, T., Batty, E.M., Fung, R., Larner-Svensson, H., Votintseva, A.A., Miller, R.R., Godwin, H., Knox, K. and Everitt, R.G. (2012) Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. *Proceedings of the National Academy of Sciences*, **109**, 4550-4555.

11) Segerman, B. (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in cellular and infection microbiology*, **2**, 116.

12) Steinberg, A.P., Lin, M. and Kussell, E. (2022) Core genes can have higher recombination rates than accessory genes within global microbial populations. *Elife*, **11**, e78533.

13) Álvarez, V.E., Quiroga, M.P., Galán, A.V., Vilacoba, E., Quiroga, C., Ramírez, M.S. and Centrón, D. (2020) Crucial role of the accessory genome in the evolutionary trajectory of Acinetobacter baumannii global clone 1. *Frontiers in microbiology*, **11**, 465735.

14) Chaves-Moreno, D., Wos-Oxley, M.L., Jáuregui, R., Medina, E., Oxley, A.P. and Pieper, D.H. (2015) Application of a novel "pan-genome"-based strategy for assigning RNAseq transcript reads to Staphylococcus aureus strains. *PloS one*, **10**, e0145861.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

15) Ozer, E.A., Allen, J.P. and Hauser, A.R. (2014) Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC genomics*, **15**, 1-17.

16) Karmakar, A., Dua, P. and Ghosh, C. (2016) Biochemical and molecular analysis of Staphylococcus aureus clinical isolates from hospitalized patients. *Canadian journal of infectious diseases and medical microbiology*, **2016**.

17) Patel, J., Cockerill, F., Bradford, P., Eliopoulos, G., Hindler, J., Jenkins, S., Lewis, S., Limbago, B., Miller, A. and Nicolau, P. (2015) M07-A10 Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard—. *Clinical Laboratory Standards Institute*, **35**.

18) Nag, S., Larsen, G., Szarvas, J., Birkedahl, L.E.K., Gulyás, G.M., Ciok, W.J., Lagermann, T.M., Tafaj, S., Bradbury, S. and Collignon, P. (2023) Whole genomes from bacteria collected at diagnostic units around the world 2020. *Scientific Data*, **10**, 628.

19) Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072-1075.

20) Antipov, D., Korobeynikov, A., McLean, J.S. and Pevzner, P.A. (2016) hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics*, **32**, 1009-1015.

21) Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, **25**, 1043-1055.

22) Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068-2069.

23) Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M. and Kubal, M. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, **9**, 1-15.

24) Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D.W. and Aarestrup, F.M. (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of clinical microbiology*, **50**, 1355-1361.

25) Jolley, K.A., Bray, J.E. and Maiden, M.C. (2018) Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. *Wellcome open research*, **3**.

26) Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D. and Kenyon, R. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic acids research*, **42**, D581-D591.

27) Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, **31**, 3691-3693.

28) Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L. and Nikolskaya, A.N. (2003) The COG database: an updated version includes eukaryotes. *BMC bioinformatics*, **4**, 1-14.

29) Rasheed, N.A. and Hussein, N.R. (2021) Staphylococcus aureus: an overview of discovery, characteristics, epidemiology, virulence factors and antimicrobial sensitivity. *European Journal of Molecular & Clinical Medicine*, **8**, 1160-1183.

30) Altaf, N. (2015) Prevalent clones of methicillin resistant Staphylococcus aureus strains in Pakistan by various typing techniques. *Annals of Pathology and Laboratory Medicine*, **2**.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13831558

31) Madzgalla, S., Syed, M., Khan, M., Rehman, S., Müller, E., Reissig, A., Ehricht, R. and Monecke, S. (2016) Molecular characterization of Staphylococcus aureus isolates causing skin and soft tissue infections in patients from Malakand, Pakistan. *European Journal of Clinical Microbiology & Infectious Diseases*, **35**, 1541-1547.

32) Shabir, S., Hardy, K.J., Abbasi, W.S., McMurray, C.L., Malik, S.A., Wattal, C. and Hawkey, P.M. (2010) Epidemiological typing of meticillin-resistant Staphylococcus aureus isolates from Pakistan and India. *Journal of medical microbiology*, **59**, 330-337.

33) Zafar, A., Stone, M., Ibrahim, S., Parveen, Z., Hasan, Z., Khan, E., Hasan, R., Wain, J. and Bamford, K. (2011) Prevalent genotypes of meticillin-resistant Staphylococcus aureus: report from Pakistan. *Journal of medical microbiology*, **60**, 56-62.

34) Baig, S., Johannesen, T.B., Overballe-Petersen, S., Larsen, J., Larsen, A.R. and Stegger, M. (2018) Novel SCCmec type XIII (9A) identified in an ST152 methicillin-resistant Staphylococcus aureus. *Infection, Genetics and Evolution*, **61**, 74-76.

35) Lim, K.T., Yeo, C.C., Suhaili, Z. and Thong, K.L. (2012) Comparison of methicillin-resistant and methicillin-sensitive Staphylococcus aureus strains isolated from a tertiary hospital in Terengganu, Malaysia. *Japanese journal of infectious diseases*, **65**, 502-509.

36) Wu, Z., Li, F., Liu, D., Xue, H. and Zhao, X. (2015) Novel type XII staphylococcal cassette chromosome mec harboring a new cassette chromosome recombinase, CcrC2. *Antimicrobial agents and chemotherapy*, **59**, 7597-7601.

37) Preeja, P.P., Kumar, S.H. and Shetty, V. (2021) Prevalence and characterization of methicillin-resistant Staphylococcus aureus from community-and hospital-associated infections: a tertiary care center study. *Antibiotics*, **10**, 197.

38) Naz, K., Ullah, N., Naz, A., Irum, S., Dar, H.A., Zaheer, T., Shahid, F. and Ali, A. (2022) The epidemiological and pangenome landscape of Staphylococcus aureus and identification of conserved novel candidate vaccine antigens. *Current Proteomics*, **19**, 114-126.