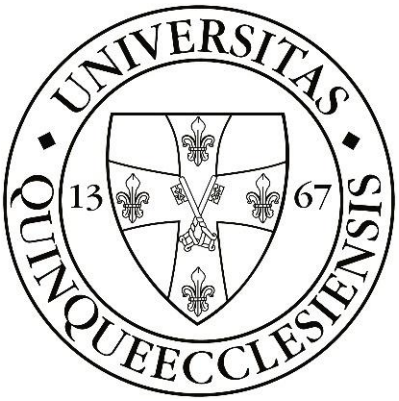
UNIVERSITY OF PÉCS

Doctoral School of Biology and Sportbiology

Deciphering the genetic architecture of methicillin-resistant *Staphylococcus aureus* clinical isolates using whole-genome sequencing

Ph.D. thesis

Naorem Romen Singh



PÉCS, 2022

UNIVERSITY OF PÉCS

Doctoral School of Biology and Sportbiology

Deciphering the genetic architecture of methicillin-resistant *Staphylococcus aureus* clinical isolates using whole-genome sequencing

Ph.D. thesis

Naorem Romen Singh

Supervisor:

Dr. habil. Csaba Fekete

Ph.D.

Supervisor Head of Doctoral School

PÉCS, 2022

Table of Contents

List of abbreviations

°C	Degrees (Celsius)
aa	Amino acid
aac6'-aph2'	6′-acetyltransferase-2′′-phosphotransferase (gentamicin resistance)
AAI	Average amino-acid identity
ANI	Average nucleotide identity
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CA-MRSA	Community-associated Methicillin-resistant *Staphylococcus aureus*
CC	Clonal complex
CFU	Colony forming units
CLSI	Clinical and Laboratory Standards Institute
coa	coagulase
COG	Cluster of Orthologous groups
CoNS	coagulase-negative *Staphylococcus aureus*
DNA	Deoxyribonucleic acid
HA-MRSA	Hospital-associated Methicillin-resistant *Staphylococcus aureus*
HGT	Horizontal Gene Transfer
IS	Insertion sequence
kb	Kilo bases
Mb	Mega bases
MGE	Mobile Genetic Element
min	Minutes
ml	Milliliter
MLST	Multilocus Sequence Typing
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSSA	Methicillin-sensitive *Staphylococcus aureus*
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
OD	Optical density
PBP	Penicillin-binding protein
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PVL	Panton-Valentine leukocidin
RFLP	Restriction Fragment Length Polymorphism
SaPI	Staphylococcal Pathogenicity Island
SCC*mec*	Staphylococcal Cassette Chromosome mec
SNP	Single Nucleotide Polymorphism
Spa	Staphylococcus Protein A
ST	Sequence Type
Tn	Transposon
TSA	Tryptic Soy Agar
TSB	Tryptic Soy Broth
WGS	Whole Genome Sequencing

1. Introduction

Staphylococcus aureus is a major human and animal pathogen, leading to cause severe hospital-associated, community-associated, and animal-associated infections (Gordon & Lowy, 2008). *S. aureus* colonies in various ecological niches within human and animal hosts (Deghorain & van Melderen, 2012) and cause a diverse range of infections ranging from skin and soft tissue infections to life-threatening infections (Feng et al., 2008; Tenover et al., 2019). The animal-associated *S. aureus* can be a potential risk of human zoonoses and a threat to human public health (Fitzgerald, 2012; Fluit, 2012). *S. aureus* acquires an arsenal of antibiotic resistance genes (ARGs) and virulence factors-encoding genes (VFGs) that are subjected to horizontal gene transfer (HGT) and recombination (Chan et al., 2011). The genomic plasticity of *S. aureus* enabled the emergence of hypervirulent and multi-drug resistant (MDR) strains and led to challenging issues in antibiotic therapy. Consequently, the morbidity and mortality rates caused by *S. aureus* infections have a substantial impact on health concerns (Denis, 2017). Thus, the prevention and control of MDR *S. aureus* infections have become the main concern in the public health sectors of the European countries (Köck et al., 2010; Smith et al., 2002). This strongly suggests that there is a continuous need to search for additional drug or vaccine targets in their genomes that would improve protection and long-lasting preventive measures and strategies. (Perumal et al., 2007).

The phenotyping and molecular typing methods of *S. aureus* are essential for delineating the occurrence of an epidemic and monitoring the transmission of the organism (Du et al., 2011). These methods have high discriminatory abilities and help to determine the relatedness among geographically diverse methicillin-resistance *S. aureus* (MRSA) (Grundmann et al., 2010; Singh et al., 2006) and are beneficial for identifying the risk factors associated with MRSA infections which support the establishment of adequate infection control programs (Zhang et al. 2012; Mistry et al. 2016). However, these techniques have certain limitations in infection control and investigating the nosocomial transmission as these techniques provide low resolution and are more time-consuming. The arrival of the Next Generation Sequencer (NGS) based-genome sequencing technology has brought with it the possibility of genome-wide applications in genome-based typing and investigations of pathogen outbreaks (Köser et al., 2012). The sequence data generated by such technologies provide the

knowledge of the population structure of *S. aureus*, allowing greater accuracy in describing and defining the different lineages, and mobilization of phages among the *S. aureus* strains and their diversity that offers insights into the evolutionary changes of lineages.

With the advancement in genome sequencing technologies, the number of bacterial genome sequences has increased rapidly and provides an excellent opportunity to accelerate the drug or vaccine targets discovery process through computational approaches such as comparative and reductive genomics, and reverse vaccinology (Rappuoli, 2000). These approaches enable an evaluation of bacterial proteins that can bind to drug molecules or induce an adaptive immune response (Muzzi et al., 2007). Reverse vaccinology optimizes the prediction of drug and vaccine target proteins for the pathogens that are challenging to culture in the laboratory.

Several comprehensive studies are addressing the *S. aureus* infections by concurrently exploring the epidemiology, phylogenetic reconstruction, genome comparison, and prophages diversity. However, most of the MRSA genome sequencing data were reported from a few countries and the global or pan-genome surveillance studies on these few data might deceive. So, performing whole-genome sequence analysis of more MRSA isolates will help understand the global surveillance and epidemiological monitoring. Although most research on *S. aureus* has been made to understand the development of antibiotic-resistant, biofilm production, and pathogenesis of *S. aureus* infection, there is a dearth of knowledge in terms of transfer of genetic elements, phage diversity, pathogenesis, antibiotic resistance, and drug targets and drug compound identification. Therefore, performing whole-genome sequence analysis may help screen our many unknown facts about MRSA strains vis-à-vis their pathogenesis.

This study is to understand and exploit the molecular basis of drug resistance, pathogenesis, niche-specific difference, and evolutionary relationship of closely related *S. aureus* clinical strains. Moreover, comparative genome analysis extends the understanding of prophage diversity and identification of potential drug and vaccine target proteins of *S. aureus* strains. This knowledge would allow the extrapolation of basic principles to improve diagnosis, infection control, and treatment strategies of staphylococcal disease.

2. Review of Literatures

Antibiotic resistance has been escalated to community health threats worldwide and is closely associated with unnecessary and excessive use of antibiotics in the hospital, agriculture, and livestock and these pathogens have the potential to mutate and render the antibiotics ineffective (Smith et al., 2002). *S. aureus* infection is more concerned because of its resistance to almost β-lactam antibiotics and multi-non-β-lactam drugs (Udo, 2013). In the European Union (EU), MRSA still poses a formidable clinical threat, with persistently high morbidity and mortality (Turner et al., 2019). The mortality rates of MRSA infections are higher than the combined mortality rates of HIV/AIDS, Parkinson's disease, emphysema, and homicide (Rossolini et al., 2014; Ventola, 2015). MRSA infections affect more than 150,000 patients annually in the EU, resulting in extra attribution of 380 million Euro in hospital charges (Köck et al., 2010). MRSA causes 10% of bacteremia and antibiotic-resistant rates rise more rapidly to 50% in several EU countries (O' Neil, 2014). The emergence of new MRSA strains associated with community and livestock infections has been reported in most of the EU countries (Köck et al., 2010; Smith et al., 2002). Thus, the prevention and control of MRSA infections have become the main concern in the public health sectors of the EU countries.

2.1 General features of *Staphylococcus aureus*

S. aureus is a gram-positive coccus bacterium that belongs to the genus *Staphylococcus*, family *Staphylococcaceae* and phylum *Firmicutes* (Foster, 1996; Liu, 2014). Morphologically, they are non-sporulating, non-motile cocci with a size ranging from 0.5-1.0μm in diameter (Foster, 1996; Liu, 2014). They are facultative anaerobe and able to grow at a temperature ranging from 7º C and 48.5º C, the pH ranging from pH 4.2-9.3, under the salt concentration of 6.5 to 15% (Chaibenjawong & Foster, 2011). There are more than eighty species and subspecies of *Staphylococci* of which *S. aureus* is often associated with human and livestock infections (Fitzgerald & Holden, 2016). It was first isolated by Sir Alexandar Ogston in 1881 and later revised to *Staphylococcus aureus* by Friedrich Julius Rosenbach in 1884 (Licitra, 2013).

Staphylococci are categorized into coagulase-positive and coagulase-negative groups based on their ability to clot blood plasma by the action of the enzyme coagulase (Cheng et al., 2010; Foster, 1996). *S. aureus* belongs to a coagulase-positive

group; however, some *S. aureus* strains have defected in coagulase production (Fairbrother, 1940; Freney et al., 1988). *S. aureus* is a catalase-positive, and this characteristic allows it to distinguish itself from other *Staphylococci* species (Kloos & Bannerman, 1994). Coagulase-positive staphylococci are generally more virulent (Foster, 1996). *S. aureus* exists as both a commensal bacterium and a human pathogen and has the capability of colonizing diverse ecological niches within its human host, including the respiratory tract, skin, blood, and nasal passages (DeLeo et al., 2009). Approximately, colonies are 40% in the skin and 30% in the nasal of a healthy person (Kluytmans et al., 1997). It causes diverse ranges of superficial soft tissue and skin infections (SSI) such as carbuncles, abscesses, styes, and impetigo to life-threatening infections such as endocarditis, necrotizing pneumonia, osteomyelitis, bacteremia, sepsis, and toxic shock syndrome (Foster, 1996; Mottola et al., 2016).

2.2 Methicillin-resistant *Staphylococcus aureus* (MRSA)

S. aureus infections were treated with benzylpenicillin (penicillin G), a β-lactam antibiotic during the 1940s but soon after the introduction of penicillin G, *S. aureus* appeared resistant to benzylpenicillin due to the acquisition of *blaZ* gene. This gene encodes for β-lactamase enzymes and is controlled by antirepressor (*blaR1*) and repressor (*blaI*) genes and is generally located at a transposon on a plasmid (Lowy, 2003). The β-lactamase enzyme cleaves the amide bond of the four-membered β-lactam ring and causes hydrolysis and inactivation of the β-lactam antibiotic **(Fig. 1)**. Now the widespread presence of penicillinase in *S. aureus* has rendered penicillin almost useless for most infections (Chambers & DeLeo, 2009).

In 1959, methicillin was developed, however, after two years of the introduction of methicillin, the first case of methicillin and all β-lactams resistant *S. aureus* strains was reported in the United Kingdom in 1961 (DeLeo et al., 2009; Otto 2012). This event has led to the emergence of MRSA strains and is recognized as a hospital-associated pathogen worldwide (DeLeo et al., 2010; Otto, 2012). Methicillin resistance in *S. aureus* is mediated by the acquisition of *mecA* located on a mobile genetic island designated staphylococcal cassette chromosome *mec* (*SCCmec*) by methicillin-susceptible *S. aureus* (Zhang et al., 2012). This gene encodes a penicillin-binding protein 2a (PBP2a), a cell wall transpeptidase (Katayama et al., 2000) that provides a low affinity to bind on β-lactam and confers resistance to most molecules of the β-lactam drug family (Pinho et al., 2001; Stapleton & Taylor, 2002) **(Fig. 1)**.

The *SCCmec* is a 21-67 kb mobile genetic element that consists of the *mecA* gene, a *ccr* gene complex, direct and inverted repeats (DR and IR), and junkyard regions (J1, the region between *ccr* and the chromosomal region flanking *SCCmec*; J2, the region between *mec* and *ccr*; J3, the region between *orfX* and *mec*) that integrates into a specific bacterial chromosomal site called *attB* at the 3' end of the *orfX* (unknown open reading frame) gene through recombination between the *attSCC* site on the circular *SCCmec* and the *attB* site (Ito et al., 2014; Katayama et al., 2000; Noto et al., 2008). This results in the integrated *SCCmec* being flanked by the *attL* site within *orfX* and the *attR* site on the chromosome. The integration and excision of the *SCCmec* element at the *orfX* integration site are mediated by genes (*ccrA, ccrB, ccrC*) on the cassette chromosome recombinase complex (*ccr*) which recognize DR or IR flanking *orfX* (Ito et al., 2009). Two *ccr* complexes have been described, one of which contains the *ccrA* and *ccrB* genes and the other containing only *ccrC* (Ito et al., 2009; Katayama et al., 2000). Overexpression of *ccrAB* has been shown to trigger the excision of *SCCmec* as a circular molecule (Noto & Archer, 2006).

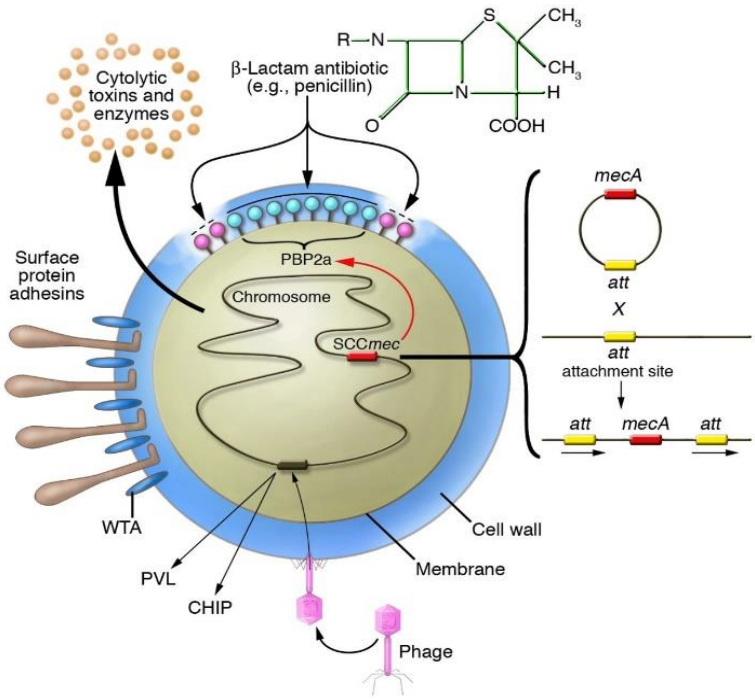


Fig. 1. *S. aureus* acquires *mecA* gene (Foster, 2004).

MRSA infections were recognized as traditionally hospital-associated (HA-MRSA), and it was discovered in 1961 that their infections have been a major threat to public health causing severe nosocomial infections worldwide (Chamchod &

Palittapongarnpim, 2019). However, in the 1980s, community-associated MRSA (CA-MRSA) has been reported among individuals living in remote communities in Western Australia with no previous history of hospitalization (Udo et al., 1993). The CA-MRSA possess novel, small *SCCmec* type IV or V or VI or VII genetic element associated with *mecA* gene and may contain with or without extra antibiotic resistance genes and these elements are easier to transfer to other *S. aureus* strains, also reported that these strains often harbor the Panton-Valentine leukocidin (*pvl*) gene that leads to SSI such as impetigo, cellulitis, folliculitis, and boils; community-associated pneumonia, bacteremia, osteomyelitis (DeLeo et al., 2009; Udo & Al-Sweih, 2013). However, HA-MRSA possesses larger *SCCmec* types I or II or III genetic elements to confer resistance to multiple different classes of antibiotics and tend to multiply slowly in culture media (Armand-Lefevre et al., 2005). These HA-MRSA strains are responsible for causing infections such as bloodstream infections, urinary tract infections, respiratory tract infections, surgical wounds, and device-associated infections (Spaulding et al., 2013). The strains associated with nosocomial infections are frequently spread into the community and unclear to distinguish between HA-MSRA and CA-MSRA. The use of *SCCmec* typing alone as a marker for CA-MRSA could lead to misclassification in the case of the EMRSA-15 clone since this clone carries *SCCmec* IV, however, this clone is the most widely known HA-MRSA (Udo, 2013). Therefore, a combination of multilocus sequence typing (MLST) or Spa typing and *SCCmec* typing has been recommended as criteria for defining CA-MRSA (Udo, 2013).

The transmission of MRSA strains from animals to humans is bidirectional (Spoor et al., 2013). Livestock-associated *S. aureus* (LA-MRSA) strains may spread to the human population through various routes such as contact with contaminated meat products, or infected farmers, butchers, and veterinary staff. Also, the contaminated effluent released from the animal farmhouses or veterinary hospitals could be another route for the transmission of *S. aureus* from animals to humans (Smith, 2015).

2.3 Epidemiology of MRSA and molecular typing methods

Microbiologists have been using phenotyping techniques such as serotyping, biotyping, bacteriophage typing, and antimicrobial susceptibility profiles to distinguish among the *S. aureus* isolates. The bacteriophage typing is poor

reproducibility and need more technical, and antimicrobial susceptibility profiling has not been very discriminatory for the analysis of nosocomial MRSA because most are resistant to many classes of antibiotic; therefore, these phenotyping techniques do not allow for the differentiation between related and unrelated of *S. aureus* isolates (Mehndiratta & Bhalla, 2012; Viau et al., 2015). Epidemiological studies of MRSA using various molecular typing techniques such as Pulse Field Gel Electrophoresis (PFGE) and Multilocus Sequence Typing (MLST), Multilocus Enzyme Electrophoresis (MLEE), *SCCmec* typing, *coa* typing, and *spa* typing have high discriminatory abilities and able to determine different MRSA clones (Goudarzi et al., 2017). Many of these established techniques are expensive and time-consuming, and the discriminatory abilities of these techniques are also different (Du et al., 2011). However, the *spa* typing method has been considered a rapid and inexpensive method for genotyping and it provides high discriminatory power than other methods (Goudarzi et al., 2017; Kareem et al., 2020). Both phenotypic and molecular typing methods have been used widely to detect and differentiate several MRSA strains, but these methods are technically demanding, labor-intensive, and time-consuming (Sabat et al., 2013). Because of that in recent times whole genome-based typing has been used as it offers an excellent resolution in global and local epidemiological investigations of pathogen outbreaks and offers further data mining activities essential for ARGs and VFGs profiling (Köser et al., 2012).

2.4 Genome organization of MRSA

The genomes of MRSA are plastic, and the genomic size varies from ~2.5 to 2.9 Mbp (Holden et al., 2013). The first whole-genome sequence (WGS) of *S. aureus* strains was done in MRSA strain N315 and vancomycin-resistant *S. aureus* strain Mu50 in 2001 (Kuroda et al., 2001). The genomic analysis of *S. aureus* strains categories the genome into three regions such as the core-genome, core-variable, and accessory genome (Lindsay, 2010). The core genome comprises about 70% conserved genes which are housekeeping genes as well as a few conserved virulence determinants, responsible for cell survival, such as genes encoding molecules involved in metabolism, DNA and RNA synthesis, and replication. The core variable genome comprises 10% genes predominantly conserved in strains sharing the same evolutionary history. The single nucleotide polymorphism (SNPs) of the core variable genome is the ultimate typing tool for the confirmation of *S. aureus* genome clonal

nature (Lindsay et al., 2006). SNPs are exploited in the technique of MLST which estimates the phylogeny from allelic variation in seven representative housekeeping genes (Enright & Spratt, 1999). Also, the difference in the repetitive sequence length results in the core genome variation. The variation in the length of short repeats that encode the Ser-Asp dipeptide repeat region R of the Clf-Sdr family and region Xr of protein A were found within the core-genome (McCarthy & Lindsay, 2010; Shopsin et al., 1999). In Sdr and CNA surface proteins, the number of 110-113 residue 'B' repeats and 180 residue B repeats, respectively can be varied from one strain to another (Foster et al., 2014). Diversity in loci encoding FnBPA/FnBPB and SdrC/SdrD/SdrE can also occur with some strains that carry only a single *fnbA* gene or missing one of the three tandemly arrayed *sdr* genes. Lastly, the accessory genome is highly variable and plays the role of the most rapid and dramatic form of genetic adaptation and are mainly mobile genetic elements (MGEs) such as chromosome cassettes, insertional sequence element (IS), plasmids, genomic or pathogenicity islands, prophages, integrative conjugative elements (ICEs) and transposons (Hacker & Carniel, 2001). The MGEs are transferred within and between the species lineage through HGT and they are mainly the genes that encode the resistance genes and virulence factors encoding genes (Lindsay et al., 2006).

The next-generation sequencer (NGS) technologies have been used to discover virulence factors encoding genes that enable it to invade and colonize a host, distinguish novel species or strains of an organism, and track disease outbreaks (Bos et al., 2019), and study the relationship between organisms. Also, this technology allowed the researchers to distinguish the essential genes of the pathogenic strains for drugs or vaccine development (Barh et al., 2011). So, the NGS based-genome sequencing technique has become a vital tool in the clinical microbiology arena for comparative genomic analysis of several other species of the Staphylococcus genus in terms of the niche adaptation, combat antibiotics, and emergence of new virulent strains in real-time (Maljkovic Berry et al., 2020; McClure et al., 2018; Raven et al., 2020; Tenover et al., 2019).

2.4.1 Antibiotic resistance genes

MRSA is not only a major threat due to their β-lactam resistance nonetheless also because of their high adaptability under selection pressure. Penicillin and methicillin resistance were gained within a short period and other non-β-lactam

antibiotic resistance determinants can easily be taken up by the integration of plasmids or transposons into their *SCCmec* element (Aires De Sousa & de Lencastre, 2004; Deurenberg et al., 2007). Vancomycin was the only antimicrobial agent that was active against all staphylococci; therefore, vancomycin has been the drug of last resort for treating infection caused by MDR-MRSA strains. Vancomycin was used for over 30 years before resistance was detected in MRSA, however, clinical strains of *S. aureus* with intermediate susceptibility to vancomycin were first reported in Japan in 1997 (Hiramatsu et al., 1997). Vancomycin resistance in *S. aureus* is maintained by retaining an original enterococcal plasmid or by a transposition of Tn1546 from the vancomycin-resistant enterococci (VRE) plasmid into a staphylococcal resident plasmid (McGuinness et al., 2017; Périchon & Courvalin, 2009). The acquisition of the *vanA* operon in vancomycin-resistance *S. aureus* (VRSA), results in alteration of peptidoglycan stem peptide synthesis, changing the terminal D-Ala-D-Ala to D-Ala-D-Lac, which has effectively reduced affinity for vancomycin (Appelbaum, 2007; Sieradzki et al., 1999; Walsh, 2003). Besides being resistant to most β-lactam antibiotics or vancomycin, MRSA is frequently associated with resistance to other classes of antibiotics. The development of such drug resistance in these pathogens is mainly due to the alteration of the drug target site, enzymatic inactivation of the antimicrobial agent, efflux pump, and sequestration of the antimicrobial agent. Other resistance mechanisms involve the acquisition of resistance determinants, position selection, and spontaneous mutation (Bitrus et al., 2017; Pantosti et al., 2007).

2.4.2 Virulence factors of *S. aureus*

MRSA was described as a versatile pathogen that encodes a wide array of virulence factors to facilitate successful adaptation and emergence of new and highly resistant and pathogenic clones (Lazarevic et al., 2011). The toxin carried by *S. aureus* consists of three major groups such as pore-forming toxins (PFTs), exfoliative toxins (ETs), and superantigens (Sag) (Götz et al., 2006; Grumann et al., 2014). These toxins contribute to a wide range of different staphylococcal infections including toxic shock syndrome (TSS), staphylococcal scalded skin syndrome (SSSS), necrotizing pneumonia, or deep-seated skin infections (Dinges et al., 2000; Jarraud et al., 1999; Ladhani, 2003; Otto, 2010). These toxins damage the host's cell membrane either by degrading inter-cellular connections or by modulating immune responses (Grumann et al., 2014).

PFTs or hemolysins consist of α-Hemolysin (Hla or α-toxin), β-Hemolysin (Hlb or β-toxin), γ-hemolysin (Hlg or leukotoxins), and δ-hemolysin (phenol-soluble modulins, PSMs) (Grumann et al., 2014). α-Hemolysin is a β-barrel forming cytotoxin that alters the cell signaling pathways including cell proliferation, inflammatory responses, cytokine secretion, and cell-cell interactions (Bhakdi & Tranum-Jensen, 1991; Bhakdi et al., 2004; Parker & Feil, 2005). This toxin is secreted by 95% of clinical *S. aureus* strains (Grumann et al., 2014; Otto, 2010). β-Hemolysin (Sphingomyelinase C or Phospholipase C) damages eukaryotic membranes by enzymatic alteration of sphingolipid content (Que & Moreillon, 2014), also responsible for the phagosomal escape of *S. aureus* and biofilm formation (Huseby et al., 2007; Katayama et al., 2013). γ-Hemolysin (Hlg) is also referred to as leukocidin and can lyse white blood cells (Spaan et al., 2015). It is required by MRSA for survival and proliferation during bacteremia (Spaan et al., 2015). *δ*-Hemolysin (PSMα1-4, PSMmec, PSMβ1-2) are hemolytic to erythrocytes, various organelles, bacterial protoplasts, and spheroplasts (Verdon et al., 2009). PSMα lyse neutrophils post-phagocytosis (Surewaard et al., 2013) and contributes to biofilm formation (DeLeo et al., Otto, 2012).

Panton-Valentine Leukocidin (PVL) is a cytotoxin that causes the destruction of leukocytes and tissue necrosis and is often associated with a specific type of pneumonia called necrotizing pneumonia (Darboe et al., 2019). PVL is a bacteriophage-encoded bicomponent leukotoxin encoded by *lukS-PV* and *lukF-PV* genes. It resides in the genomes of several icosahedral- or elongated-head-shape temperate bacteriophages including ϕSa2958, ϕSa2MW, ϕPVL, ϕ108PVL, ϕSLT, and ϕSa2USA. The PVL is observed in a few strains of *S. aureus* strains, and mostly in CA-MRSA strains (Boakes et al., 2011). LukED and LukAB/GH enhances the survival of *S. aureus* by playing role in escaping the *S. aureus* from phagocytes and neutrophils, also targeting monocytes, dendritic cells, and leukocytes (DuMont et al., 2013; Melehani et al., 2015).

ETs are serine proteases that recognize and hydrolyze desmosome cadherins in the skin's superficial layers (Mariutti et al., 2017). The principal ETs are known so far are ETA, ETB, ETC, and ETD. ETA and ETB are the principal isoforms of exotoxins involved in human skin damage, while ETC was isolated from a horse infection and had no association with human disease (Mariutti et al., 2017). ETA is codified by the *eta* gene on chromosomal DNA, carried on the genome on a temperate phage, and

ETB by the *etb* gene on a large plasmid (Que & Moreillon, 2014) while ETD is codified by the *etd* gene which is located chromosomally on a pathogenicity island of *S. aureus* clinical sample (Nishifuji et al., 2008). Approximately 5% of *S. aureus* strains carried ETs, ETA is most prevalent in Europe, Africa, and America while ETB is more common in Japan (Ladhani, 2001).

Superantigens were known as staphylococcal enterotoxins (SEs) since they cause symptoms typical of *S. aureus* food poisoning such as vomiting and diarrhea (Grumann et al., 2014). Based on their antigenicity, SEs are of 23 different SEs and enterotoxin-like (SEls) types such as SEA to SEE, SEG to SEJ, SEL to SEQ, and SER to SET, and SEIs namely SEIK to SEIQ, SEIU to SEIX (Hu & Nakane, 2014). The SEs trigger T-cell activation and proliferation; their mode of action may involve the activation of cytokine release and cell death *via* apoptosis and potentially lethal TSS (Lin et al., 2010). The genes for SEs and SEls are located on mobile genetic elements, including plasmids, transposons, prophages, *S. aureus* pathogenicity islands (SaPI), variable genomic region vSaβ, or next to the SCC elements. Only the staphylococcal gene cluster *egc* is organized as an operon. These toxins are mobile elements, thus horizontal transfer between strains is more frequent (Hiramatsu et al., 2001).

2.4.3 Biofilm as a virulence factor

S. aureus can anchor on epithelial surfaces and form multicellular communities known as biofilms (Goudarzi et al., 2017; Strandén et al., 2003). Chronic infections are associated with a biofilm formation in which the *S. aureus* is attached and persists on host tissue especially in bone and heart valves or on implanted materials such as catheters, prosthetic joints, and pacemakers (Barrett & Atkins, 2014; Maiti et al., 2014). This results in causing chronic wound infections, chronic urinary tract infections (UTI), osteomyelitis, cystic fibrosis pneumonia, cystic fibrosis pneumonia, and endocarditis (Stefanaki et al., 2017). Biofilm is a complex 3D structure of a sessile microbial community covered by an exopolysaccharide glycocalyx matrix (Otto, 2008). A major constituent of the biofilm matrix is polysaccharide intercellular adhesin (PIA), which is composed of β-1,6-linked N-acetylglucosamine with partially deacetylated residues (Mack et al., 1996; Vuong et al., 2004). The *icaADBC* gene cluster is responsible for PIA synthesis (Arciola et al., 2015; Chaieb et al., 2005; Hoang et al., 2019). The PIA plays an important role in the structural integrity of biofilms *in vitro* and *in vivo* conditions. It was reported that the co-existence of *icaA*

and *icaD* increases N-acetylglucosaminyltransferase activity and slime production (Arciola et al., 2001; Arciola et al., 2006). It is also reported that the administration of antibiotics to resistant bacteria promotes biofilm formation by inducing the expression of *ica*A-dependent polysaccharide intracellular adhesin (PIA) production (Chadha, 2014). However, *S. aureus* strains have capable of forming biofilm in the absence of *ica* operon (Foster et al., 2014). In such a PIA-independent case, the number of surface proteins and eDNA involve in biofilm formation (Boles et al., 2010; O'Neill et al., 2007, 2008; Rohde et al., 2007) The eDNA being negatively charged can anchor cells to a surface, to host factors, and to each other during attachment (Mann et al., 2009). Autolysin (Atl) is a wall-anchored protein of *S. aureus* and causes initial attachment to the surface that can be cleaved into amidase and glucosaminidase resulting in cell lysis, the release of eDNA, and cell accumulation (Archer et al., 2011). The fibronectin-binding proteins (FnBPs) can also arbitrate biofilm formation and maturation through an essential role by the major autolysin (Atl) and SigB regulation (Archer et al., 2011). In biofilm formation, α-hemolysin facilitates cell-cell interactions, β-hemolysin causes hemolysis, and lyse lymphocytes play a stimulative role in the biofilm formation by covalently cross-linking to itself in the occurrence of DNA in a matrix of staphylococcal biofilms. *S. aureus* strains of which *agr* classes are *agr* II and *agr* III are high and medium biofilm formers due to having defective and inactive *agr*, respectively. Non-defective and active *agr* is present in *agr* I and *agr* IV strains that are weak biofilm producers (Cafiso et al., 2007).

2.4.4 Mobile Genetic Elements (Mobilome)

The accessory genome of *S. aureus* has a wide range of Mobile Genetic Elements (MGEs): these are DNA segments that mediate the passage of DNA within or between bacteria. These consist of *SCCmec* elements discussed earlier, plasmids, bacteriophages, transposons (Tn), insertion sequences (IS), staphylococcal pathogenicity islands (SaPIs), and integrative conjugative elements (ICEs) (Lindsay, 2010). The acquisition of such MGEs through HGT allows the *S. aureus* strains to lead to the emergence of new hypervirulent strains or MDR strains that causes new clinical challenges (Lindsay, 2010; Lindsay & Holden, 2006). These MGEs play a crucial role in dynamic contribution to evolution and pathogenesis, also responsible for clinically important phenotypic differences among these strains (Lindsay et al., 2006).

2.4.5 Plasmids

Plasmids are small extrachromosomal circular double-stranded DNA molecules and can self-replicate (Novick, 1989). They can transfer from one bacterial cell to another through a process called conjugation (McCarthy & Lindsay, 2012). MRSA genome contains at least one or more plasmids that confer resistance to antibiotics and heavy metals (Lindsay, 2010). The plasmids are of three families based on the plasmid size and conjugation ability namely, class I, II, and III (Młynarczyk et al., 1998). Class I plasmids have 1.3- 4.6 kb size with 10-55 copies number per cell and frequently carry resistance determinants or are cryptic (*e g.* pT181, pC194, pSN2, and pE194) and some can be integrated into the chromosome including mobile genetic elements on the chromosome; class II plasmids have 15-46 kb size with 4-6 copies number per cell and involved in penicillinase and aminoglycoside/trimethoprim resistance and also encode genes for resistance to antiseptics and heavy metals such as mercury or arsenate (*e.g.* pSK 1 and pIP630), and class III plasmids have size ranges from 30-60 kb. Class III plasmids carry a determinant of transfer (*tra*) by conjugative transfer of the plasmid between isolates at a low frequency. These plasmids include glycoside-resistance plasmids (pGO1, pG0400, and pCRG1600) and usually possess one or two transposons and many copies of insertion sequences (IS) (Młynarczyk et al., 1998). *Staphylococcal* plasmids, pRW001, are able to encode exfoliative toxin B (ETB) and bacteriocin immunity (Bukowski et al., 2010). Sequencing of *S. aureus* plasmid revealed novel resistance genes, such as the apramycin (*apmA*) and streptogramin A (*vgaC*) genes (McCarthy & Lindsay, 2012). Most plasmids encode resistance genes or encode toxin genes while some plasmids in which their functions are still unknown, and plasmids are termed as cryptic plasmids.

2.4.6 Transposable elements

Transposable elements, jumping gene that moves within a chromosome or between chromosomes. *S. aureus* transposons are integrated into multiple copies into the chromosome or MGEs (plasmid & *SCCmec* elements) due to their relatively small sizes. They encode a transposase gene whose product catalyzes the excision, replication, and integration of the element and can carry resistance genes, for example, Tn552 carrying *bla* for penicillinase, and Tn554 carrying resistance to erythromycin, spectinomycin, and macrolide–lincosamide (Młynarczyk et al., 1998). A small (<2.5 kb usually) transposable element that only carries genes required for transposition is

called insertion sequences (IS). Although IS harbor any resistance or virulence genes, they are responsible for the recombination and stabilization of some resistance genes and can inactivate several genes either by direct insertion or by the polar effect on nearby gene transcription (Needham et al., 1995). Larger transposons (>18 kbp) with single copies are relatively rare, but they can encode tetracycline, trimethoprim, vancomycin, or aminoglycoside resistance genes (Malachowa & Deleo, 2010).

2.4.7 Staphylococcal pathogenicity islands

Staphylococcal pathogenicity islands (SaPIs) are 12 to 27 kb mobile pathogenicity islands that contain highly conserved core genes encoding transcriptional regulatory proteins, an integrase (*int*) that recognizes the integration site on the chromosome, a terminase and a replication initiation protein, and various superantigens encoding genes such as *seb, sec, sek, sel* and *tst-1*, implicated in toxic shock and food poisoning are mobilized by helper phages (Malachowa & Deleo, 2010). They reside quiescently at specific chromosomal attachment (*att*) sites under the control of their master repressors and are packaged in phage-like particles. Gene transfer between *S. aureus* strains is higher than between different Staphylococcal species due to phage receptor specificity and the restriction barriers (Lindsay & Holden, 2006; Thomas & Nielsen, 2005). It raises intra-strain and inter-strain exchange frequency of virulence factor-encoding genes or resistance genes among the *Staphylococcus* species and enables them to establish a new lifestyle by adapting to diverse hostile environments and dynamic contribution to evolution and pathogenesis (ben Zakour et al., 2008; Novick et al., 2010).

2.4.8 Bacteriophages

The bacteriophage that infects *S. aureus* strains is a dsDNA phage and comes under the *Caudovirales* order (Kwan et al., 2005). Bacteriophage genome inserted and integrated into the circular bacterial DNA chromosome or existing as an extrachromosomal plasmid known as a prophage (van Wamel et al., 2006). Prophage constitutes 10–20% of a bacterium's genome and protects the cell from lytic infection or provokes cell lysis through prophage induction (Casjens, 2003). Acquisition of prophages or phage-related genomic islands determines the diversity of the *S. aureus* species and contributes to a dramatic form of genetic adaptation to various host conditions (Hacker & Carniel, 2001). Based on the genome sequence and sizes of the *S. aureus* phages, phages can be grouped into three classes among *Caudovirales* order,

viz., Podoviridae family belongs to class I with the smallest genome (<20 kb), *Siphoviridae* family belongs to class II showing intermediate genome sizes (39-125 kb), and *Myoviridae* family belongs to class III with largest genome size (>125 kb) (Deghorain & van Melderen, 2012; Kwan et al., 2005). The genome of the *Siphoviridae* family is composed of six functional modules *viz*., lysogeny, DNA replication, packaging, head, tail, and lysis (Xia & Wolz, 2014). Prophage can be switched from a lysogenic state to a lytic state in response to the metabolic state or environmental stresses of the host (Fortier & Sekulovic, 2013). The expression of a specific phage repressor gene (*cI*) inhibits the transcription of the genes required for the lytic cycle and the prophage becomes quiescent. The CI repressor also inhibits the integration of other phage genomes of the same group and confers immunity to superinfection (Labrie et al., 2010). Phages or prophage-like elements contribute to the horizontal transfer of pathogenicity islands that carry VFGs such as PVL encodes *lukFS-PV* genes, the immune evasion cluster (IEC) associated with human specificity (*chp, sak,* and *scn*), exfoliative toxins (*eta and etb*), and enterotoxins (*sea, see, seg, sek,* and *sep*) (Baba et al., 2002; Botka et al., 2015; McCarthy, Breathnach, et al., 2012; McCarthy, Witney, et al., 2012; Novick et al., 2010). It is reported that phages 80α and 80 can mobilize a variety of superantigen-encoding pathogenicity islands SaPI1 and SaPI2, respectively (Ruzin et al., 2001; Tallent et al., 2007), and *pvl* genes are encoded by prophages phiPVL and phiSLT (Narita et al., 2001). The mobilization of phages among the *S. aureus* strains has increased the frequency of intra-strain and inter-strain exchange of VFGs or resistance genes that helps to adapt to diverse hostile environments and contributes to pathogenesis and evolution (ben Zakour et al., 2008; Novick et al., 2010; Ruzin et al., 2001).

2.5 Subtractive genomics-based and reverse vaccinology approach

MRSA is one among other MDR bacterial pathogens (Hiramatsu et al., 2014). The development of alternatives to antibiotics for the treatment and prevention of staphylococcal infections is of great concern. So, the identification of novel drug targets against the MDR-MRSA is the only possible way to address the challenge issues of antibiotic treatments. Novel drug target identification through *in vitro* or *in vivo* approaches is very expensive and time-consuming which are eventually being switched by cost-effective and time-saving automated modern computational methods, called reverse vaccinology (Rinaudo et al., 2009). The introduction of NGS

technologies provides the whole genome sequencing data of various pathogenic agents to the public. Further, the arrival of the post-genomic era has brought with it the possibility of genome-wide application of a rational new drug target and vaccine candidate selection methodology. Reverse vaccinology (RV) is used to identify and predict drug targets for several pathogenic bacteria from the genomes or proteomes by subtractive and comparative proteomics in-silico approaches (Barh et al., 2011; Pizza et al., 2000). The principle behind these approaches follows some criteria such as the identify drug target proteins should be essential for the survival and common in several strains of *S. aureus* as well as non-homologous to human genes or proteins, surface-exposed, and able to be recognized by the immune system, high adhesin capacities that involve in pathogenicity and invasion, high antigenicity, and not more than one transmembrane helix (Barh et al., 2011; Rappuoli, 2000)). The RV and subtractive genomic approaches were used to identify targets in various human pathogens (Kumar et al., 2017). A pharmaceutical company like Novartis uses such approaches and could identify more vaccine candidates in a few months than had been discovered during the last 40 years (Pizza et al., 2000). The major discovery that was made by RV is the development of a serogroup B *Neisseria meningitidis* (MenB) vaccine. This approach has been successfully applied in Group B Streptococcus (GBS), *Bacillus anthracis*, *Streptococcus pneumonia*, *Chlamydia pneumoniae*, *Porphyromonas gingivalis*, *Edwardsiella tarda*, and *Mycobacterium tuberculosis* (Motin & Torres, 2009). During Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV 2) or COVID19 pandemic, the technology was applied to predict the COVID19 vaccine candidate for example SARS-CoV-2 spike (S) glycoprotein. The identified S protein is being used as a vaccine candidate, including the Pfizer and Moderna mRNA vaccines (Ong et al., 2021). Nevertheless, RV can enable us to systemically classify the potential protective antigens, thereby helping to improve existing vaccines and develop efficient preparations virtually against any pathogen that has had its genome sequence determined. With the number of sequenced genomes progressively increasing every year, this approach, combined with recent advances in bioinformatics, comparative and functional genomics, and proteomics, will be the method of choice for vaccinology studies at the beginning of the 21st century. The main limitation of this approach is the lack of a high-throughput system to estimate the protective immunity of selected candidates.

3 Aims of the study

Characterization of methicillin-resistant *S. aureus* (MRSA) clinical isolates through genomic approach, and an in-silico identification of potential drug and vaccine candidates against methicillin-resistant *S. aureus*.

Research activities were carried out to fulfill the following key scientific objectives:

1. Characterization of MRSA clinical isolates through the polyphasic (phenotypic and genotypic) approach.
2. Genome-wide comparison of MRSA clinical isolates.
3. Comparative analysis of prophages carried by human and animal-associated *S. aureus* strains spreading across the European regions.
4. Identification of potential drug targets and vaccine candidates against MRSA strains through subtractive genomics-based and reverse vaccinology approaches.

4 Materials and Methods

4.1 Collection of the isolates

In this study, 35 *S. aureus* strains were collected from the Department of Medical Microbiology and Immunology Laboratory, Medical School, University of Pecs, Hungary. The Hungarian *S. aureus* strains (60%) were previously recovered from wounds (31.42%), blood (8.57%), tracheas (5.71%), ears (2.85%), lungs (2.85%), nostrils (2.85%), skins (2.85%) and throats (2.85%) while the German *S. aureus* strains (40%) were recovered from body sites without documentation.

4.2 Biochemical tests of the isolates

The isolates were identified as staphylococcal strains based on colony morphology on Nutrient agar, Blood Agar and Mannitol Salt Agar, Gram's stain, and different biochemical tests (Bergey & Holt, 1994). The isolates were tested for catalase, coagulase, urease, DNase production, and mannitol fermentation test (Collee et al., 1996).

4.3 Antibiotic susceptibility test

The 35 *S. aureus* clinical strains were screened for MRSA using BBL™ CHROMagar™ MRSA II media (BD). Susceptibility of *S. aureus* strains to oxacillin (1μg), cefoxitin (30μg), and erythromycin (15μg) were determined using the disk diffusion method according to Clinical and Laboratory Standards Institute (CLSI) guidance (CLSI, 2014). The entire antimicrobial susceptibility test (AST) was repeated three times using the *S. aureus* ATCC25923 and ATCC700698 as MRSA negative and positive controls, respectively. The diameter zone of inhibition was measured in millimeters (mm).

4.4 Biofilm formation assay

Biofilm formation was performed as previously described (Rahimi et al., 2016) with some modifications. Briefly, *S. aureus* strains were cultured overnight at 37º C in tryptic soy broth (TSB) (BD, Germany) containing 0.25% (w/v) glucose. The cell density was adjusted to a final concentration of 10^6 CFU/ml in TSB supplemented with 0.25% (w/v) glucose. Cell suspensions (200 µl) were loaded into 96-well round-bottomed microtiter plate (Sarstedt, Germany), and incubated at 37º C for 18 h without shaking. Cells were washed three times with 200 µl sterile PBS (pH 7.2), dried at room temperature, and fixed with methanol (99% v/v). The dried biofilm was stained with 200 µl of 0.16% (w/v) crystal violet for 15 minutes. To remove the unbound dye,

biofilms were washed three times with PBS and air-dried. Finally, the biofilm-bound dye was solubilized with 200 µl of 95% (v/v) ethanol, and absorbance was measured at 540 nm wavelength using a Multiskan Ex microtiter plate reader (Thermo Electron Corporation, USA) in a flat-bottom 96-well plate (Costar 3599; Corning; USA). Experiments were performed in triplicates with *S. aureus* ATCC25923 as a biofilm-positive control strain.

4.5 Molecular identification and genotyping

4.5.1 Genomic DNA extraction

The DNA of the 35 strains was extracted from the overnight culture of *S. aureus* using QIAamp DNA Mini Kit (Qiagen GmbH, Hilden, Germany). The extracted DNA concentration was assayed by the Nanodrop-2000 spectrophotometer.

4.5.2 Detection of *S. aureus* species-specific sequence, *mecA*, and *pvl* genes

Genomic DNA was used for the detection of *S. aureus* species-specific sequence (Martineau et al., 1998), *mecA* (Strommenger et al., 2003), and the *pvl* toxin (Hisata et al. 2005; Karahan et al. 2007) genes. *S. aureus* ATCC25923, ATCC700698, and ATCC700699 strains were used as reference strains for *mecA* negative and positive controls, respectively.

4.5.3 *SCCmec* typing

Multiplex PCR typing of *SCCmec* gene was performed on *mecA*-positive *S. aureus* strains using primers as described previously (Zhang et al., 2005). The results were validated using a simplex PCR reaction for each primer set. The MRSA isolates with unexpected fragments or lacking fragments by both PCR methods were defined as non-typeable (NT).

4.5.4 *coa* gene typing

Among *mecA*-positive *S. aureus* isolates, *S. aureus* isolates were selected based on antibiotic susceptibility profiles, and genotyping of *coa* gene polymorphism was performed using PCR amplification of the *coa* gene and confirmed by gel electrophoresis. For PCR-RFLP, the *coa* gene amplicons were digested with *HaeIII* (Fermentas, USA) restriction enzyme (Khoshkharam-Roodmajani et al., 2014) and a heat-map with dendrogram was generated from the restriction banding pattern using Morpheus web-based program (Morpheus) using the Euclidean distance feature.

4.5.5 *spa* gene typing and sequencing of polymorphic *spa* gene

Polymorphism of the *spa* gene was detected based on a previously described primer set (Harmsen et al., 2003). The PCR products were purified using the ZR-96 DNA Clean-up Kit (Zymo Research, USA). Concentration was determined by Qubit 3.0 and sequencing reactions were performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Sequencing reactions were run on ABI PRISM 310 Genetic Analyzer (Applied Biosystems, USA). The *spa* sequence types were assigned using spaTyper (Spa Type Finder/Identifier) and confirmed by using a *spa* database (Ridom SpaServer - SpaTypes) in DNAGear software (Al-Tam et al., 2012).

4.5.6 Determination of numerical index of discrimination

The discriminatory power (DP) of the genotyping method was determined based on the index described by Hunter & Gaston (Hunter & Gaston, 1988). The DPs of *coa* and *spa* typing were calculated based on Simpson's index using the online tool DP calculator (http://insilico.ehu.es/mini_tools/discriminatory_power/).

4.5.7 Identification of biofilm-encoding genes

Biofilm-forming genes, such as *icaA-D* (encoding intracellular adhesion proteins A-D), and *icaR* (encoding intracellular adhesion protein R); *fnbA* and *fnbB* (encoding fibronectin-binding protein A and B); *cna* (encoding collagen-binding protein); *clfA* and *clfB* (encoding clumping factors A and B); and *ebpS* (encoding elastin-binding protein) were detected using the primer sets listed in **Table S3**. Simplex PCRs were performed using DreamTaq PCR Master Mix according to the manufacturer's recommendation (Thermo Fisher Scientific, USA) in a Veriti™ 96-Well Thermal Cycler (Applied Biosystem, USA) as follows: 96 ºC for 3 min; 96 ºC for 30 sec, 54 ºC for 30 sec, 72 ºC for 1 min repeated for 35 cycles; final extension was performed at 72 ºC for 7 min. The amplified products were electrophorized on 2% (w/v) agarose gel, stained using 0.5 µg/ml ethidium bromide solution, and captured using the FluroChem Q system (ProteinSimple™, USA).

4.6 Data setting for polyphasic approach

Individual result of the applied techniques was converted into the unweighted binary code (0, 1), and Jacquard’s similarity index was generated and visualized according to the Neighbour-Joining (NJ) clustering method using Past 3. x (Hammer

et al., 2001). Besides the binary data was used to perform a logistic Principal Component Analysis (PCA) in R software (RStudio Team, 2020).

4.7 Whole-genome sequencing

4.7.1 Genomic DNA extraction

The genomic DNA was extracted using the GenElute™ Bacterial Genomic DNA Kit (Sigma, USA) following the manufacturer's recommendation. The extracted DNA samples were quantified in a Qubit 3.0 fluorometer (Invitrogen, USA) using dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific Inc. USA) and subsequently, DNA quality was visualized by agarose gel electrophoresis.

4.7.2 Library preparation and sequencing

Genomic libraries were prepared by using the NEB Next Fast DNA Fragmentation and Library Preparation Kit, developed for Ion Torrent (New England Biolabs, USA) and used according to the 200 bp protocol. After chemical fragmentation, DNA size selection was performed on precast 2% E-Gel Size Select Gel (Thermo Fisher Scientific Inc. USA). The quality of the libraries was verified using Agilent high sensitivity DNA assay kit (Agilent Technologies Inc. USA) in Agilent 2100 Bioanalyzer System (Agilent Technologies Inc. USA). For the template preparation, Ion PGM Hi-Q View OT2 Kit was used (Thermo Fisher Scientific Inc. USA). The template positive beads were loaded on Ion 316v2 Chip and sequenced using Ion PGM Hi-Q View Sequencing Kit on Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher Scientific Inc. USA).

4.7.3 Genome assembly, and annotation

In-silico trimming of adapter and barcode sequences and data analysis were performed using Torrent Suite 5.4.0 (Thermo Fisher Scientific Inc., USA) and the trimmed paired-end reads were assembled by *de novo* assembler SPAdes 3.7.1 software with 21, 33, 55, 77, 99, 127 k-mer values (Nurk et al., 2013). For identifying the closely related strains, the genome assemblies were analyzed by the kmerFinder 3.1 (Larsen et al., 2014). The genome assembly was aligned against the closely related genome defined by kmerFinder 3.1 for the contigs rearrangement using the 'Move Contigs' algorithm in Mauve 2.4.0 (Darling et al., 2010) and further, scaffolds were generated from the arranged contigs of the genome with genome of closely related strains predicted by kmerFinder 3.1 as a guide for alignment using the reference-based scaffolder MeDuSa (Bosi et al., 2015). Gene annotation of the genome assemblies was

performed *via* the fully automated RAST (Rapid Annotation using Subsystem Technology) (Aziz et al., 2008) and NCBI prokaryotic genome annotation pipelines (Tatusova et al., 2016).

4.7.4 In-silico characterization of genome assemblies

In-silico epidemiologic characterization of genome assemblies was performed using SCCmecFinder-1.2 for the identification of *SCCmec* types, spaTyper 1.0 for *spa* type, and MLST 1.8 for Multilocus Sequence Type in a web-based server provided by the Center for Genomic Epidemiology (CGE Server). In-silico *agr* (accessory gene regulator)- typing was performed using the primers described by Shopsin et al. in in-silico PCR amplification tools (Bikandi et al., 2004).

The genome assemblies were screened for plasmid replicon (*rep*) genes using PlasmidFinder 2.1 (Carattoli et al., 2014) with default parameters. The identified nonaligned contig or scaffold associated with plasmid sequences were extracted and used for the identification of full-length plasmid regions using PLSDB (Plasmid Database) version-2020-03-04 with search strategy Mash screen, and the default values were a maximum P-value of 0.1 and a minimum identity of 0.99 (Galata et al., 2019). Identified plasmids were compared with the closest reference plasmids using Easyfig version 2.2.3 (Sullivan et al., 2011).

In-silico prophage signals mining was performed using PHASTER (PHAge Search Tool Enhanced Release) (Arndt et al., 2016). Prophage sequences with PHASTER score $\leq$ 70 are considered incomplete, the score between 70-90 are regarded as questionable, while the score $\geq$ 90 are considered intact/complete prophages. The identified intact prophages were classified for their lifestyles using PHACTS (Phage Classification Tool Set) (McNair et al., 2012).

In-silico mining of candidate ARGs and VFGs were performed using CARD (Comprehensive Antibiotic Resistance Database) version 3.0.8 in RGI (Resistance Gene Identifier) version 5.1.0 platform (Alcock et al., 2020) and ResFinder 4.1 server (https://cge.cbs.dtu.dk/services/ResFinder-4.1/), and a comprehensive set of *S. aureus* VFGs was analyzed using VFDB (Virulence Factor Database) in VFanalyzer (Liu et al., 2019) and the PATRIC tool version 3.6.3 (Wattam et al., 2014). Further, heatmap and hierarchical clustering were generated to visualize the presence and absence of VFGS and ARGs in *S. aureus* strains using Morpheus web-based program (Morpheus). Secondary metabolite biosynthesis gene clusters and the detection of

genes encoding bacteriocins were analyzed using antiSMASH 5.0 (Blin et al., 2019) and BAGEL4 (van Heel et al., 2018). The prediction of chromosomal genomic islands was predicted by using IslandViewer 4 (Bertelli et al., 2017).

4.7.5 Comparative genome analysis

The ANI (average nucleotide identity) mean matrices based on BLASTn, and core genome-based AAI (average amino acid identity) mean matrices were generated with EDGAR 2.0 interface (Blom et al., 2016). The pairwise comparisons between the genomes of *S. aureus* isolates and their nearest reference genomes were conducted using GBDP (Genome BLAST Distance Phylogeny) under the algorithm trimming and distance formula d5 and calculated each distance with 100 replicates (Meier-Kolthoff & Göker, 2019). A circular genome plot of *S. aureus* isolates and their reference strains was generated with BioCircos assembled in EDGAR 2.0 (Blom et al., 2016). The functional annotation was performed using EggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) mapper 5.0 database (Huerta-Cepas et al., 2019) and RAST server-based SEED viewer (Overbeek et al., 2014).

4.7.6 Pan-genome, core genome, and singletons analysis

The pan-genome, core-genome, and singletons were calculated using six study genomes of *S. aureus* isolates (SA G5, SA G6, SA G8, SA H27, SA H29, and SA H32) in EDGAR 2.0 software framework (Blom et al., 2016). This pan-genome analysis was extended using the study genomes coupled with nine reference genomes of *S. aureus* strains such as *S. aureus* CA-347 (CP006044.1), *S. aureus* subsp. *aureus* ST228 (HE579071.1), *S. aureus* subsp. *aureus* JH9 (CP000703.1), *S. aureus* subsp. *aureus* str. Newman (AP009351.1), *S. aureus* subsp. *aureus* HO 5096 0412 (HE681097.1), *S. aureus* subsp. *aureus* NCTC 8325 (CP000253.1), *S. aureus* subsp. *aureus* Mu50 (BA000017.1), *S. aureus* subsp. *aureus* strain MRSA252 (BX571856.1), and *S. aureus* subsp. *aureus* DSM 20231 (CP011526.1). The core-genome was analyzed in the genomes set using reciprocal best BLAST hits of all CDS using the EDGAR version 2.0 software framework (Tettelin et al., 2008). The singletons were calculated for the contig of a strain by comparing them to the CDS of a set of contigs in EDGAR. The CDS that has no match with SRV (Score Ratio Value Plots) higher or equal to the master cut-off in any of the contigs were considered as singletons. The development of pan-genome and core-genome sizes was analyzed

using the core/pan development feature and the pan *vs.* core development plot was generated in EDGAR. The Rcp (ratio of core-genome to that of pan-genome) was calculated (Ghatak et al., 2016). Then, genomic subsets, including the number of core-genome and singletons in the gene pool, were extracted, and the flowerplot was drawn using *in-house* R scripts.

4.7.7 Phylogenetic analysis

The genome assemblies of the isolates were used for a whole genome-based phylogeny analysis using TYGS (Type/Strain Genome Server) (Meier-Kolthoff & Göker, 2019) engaging with genomes of closely related strains of *S. aureus*. The phylogenomic trees were reconstructed using FastME 2.1.6.1 (Lefort et al., 2015) from the GBDP (Genome BLAST Distance Phylogeny) distances calculated from genome sequences under the algorithm 'coverage' and distance formula d5 (Meier-Kolthoff et al., 2013). The trees were rooted at the midpoint (Farris, 1972); branch supports were inferred from 100 pseudo-bootstrap replicates and visualized with Interactive Tool Of Life v4 (iTOL) (Letunic & Bork, 2019). The core SNPs of genome sequences were extracted using Panseq (Laing et al., 2010) and the phylogenetic tree was constructed using the PhyML+SMS module in NGPhylogeny.fr (Lemoine et al., 2019) to select the best evolutionary model, further the tree was annotated in iTOL (Letunic & Bork, 2019).

4.8 Data collection and Identification of prophages

A total of 60 whole genomes of *S. aureus* strains reported to cause human and animal infections across the European regions were used in this study. Of these 60 whole genome sequences of *S. aureus* strains, 54 were retrieved from the NCBI database and additional six genome assemblies of *S. aureus* were from this current study. The *S. aureus* strains used in this study originated from Austria (n=7), Denmark (n=5), France (n=12), Germany (n=11), Hungary (n=3), Italy (n=9), Netherlands (n=11), and Spain (n=2). The genome sequences were analyzed for *SCCmec* types (Kaya et al., 2018), and MLST (Larsen et al., 2012) using a web-based server provided by the Center for Genomic Epidemiology. The prophage sequences or phages associated with these genomes were analyzed for their diversity based on the geographic location and nature of *S. aureus* infected hosts. The details of the whole genomes used in this study are presented in **Table S7.**

4.8.1 General genomic features of the putative prophages

In-silico prophage signals mining from 60 *S. aureus* genomes was performed using the same method mentioned in section 4.7.4 (third paragraph). Further, the intact/complete prophage genomes were re-annotated using prokka 1.14 (Seemann, 2014).

4.8.2 Sequence clustering and phylogenetic relationship of the prophages

A total of 65 intact prophage sequences of *S. aureus* strains were identified by PHASTER. The 65 intact prophage nucleotide sequences were subjected to Multiple sequence Alignment using Fast Fourier Transform (MAFFT) version v7.475 (Katoh et al., 2018). Further, the aligned sequences of intact prophage nucleotide sequences were run on SplitsTree4 software (Huson & Bryant, 2006) to generate the hierarchical clusters and displayed as a phenogram using the BioNJ algorithm (Gascuel, 1997).

4.8.3 Comparative genomic analyses of *S. aureus* prophages

The identified intact prophage sequences were in-silico analyzed for the identification of ARGs and VFGs using CARD (Alcock et al., 2020) and VirulenceFinder-2.0 Server (Joensen et al., 2014), respectively. The heatmap was generated to illustrate the presence or absence of VFGs using Morpheus (Morpheus). The intact prophage sequences were used for a pan-genome comparison using the TBLASTX and prophage phiH14-1 as a seed genome in the Gview server (https://server.gview.ca/). Furthermore, the prophage sequences belonging to each cluster or clade were analyzed for core and accessory genomes using Spine and AGEnt version 0.3.1 webserver (Ozer et al., 2014). The number of core and accessory genomes of prophages in the gene pool of each cluster or clade was extracted, and a flowerplot was generated using plotrix in RStudio 1.3 (RStudio_Team, 2020). The intact prophage sequences that comprised each cluster were aligned using Easyfig version 2.2.3 (Sullivan et al., 2011). Easyfig alignments were performed on selected groups of prophages based on their clusters to show regions of sequence identity and their closest phages (phiNM-3, phiStauST398-3, and phi2958PVL) defined by PHASTER.

4.9 Identification and characterization of potential vaccine and drug target candidates by reverse vaccinology

4.9.1 Retrieval of MRSA genome sequences

The genome sequences of 16 MRSA strains were used for the identification of potential vaccine and drug targets. Among the 16 MRSA genome sequences, 6 MRSA

genomes (SA G5, SA G6, SA G8, SA H27, SA H29, and SA H32) were from the current study, and 10 MRSA genomes (NCTC 8325, CA-347, ST228, JH9, Newman, HO 50960412, Mu50 DNA, MRSA252, H-EMRSA-15, and DSM 20231) were retrieved from NCBI database (*http://www.ncbi.nlm.nih.gov/genbank/*). The chosen 16 MRSA strains belonged to CC5, CC8, CC22, CC30, and CC45 **(Table S5)**. It was reported that these CCs were the most global dominant MRSA clones or lineages (Chatterjee & Otto, 2013; Monecke et al., 2011) with the high capacity to carry a large set of ARGs and to cause nosocomial infections (Chua et al., 2014).

4.9.2 Prediction of core-genome

The genomes were re-annotated using RAST (Rapid Annotation using Subsystem Technology) platform (Aziz et al., 2013) to avoid unexpected and inappropriate gene interpretation results. The identification of conserved proteins (core-proteome) between the sixteen genomes was analyzed in the genomes set using reciprocal best BLAST hits of all CDS in EDGAR (version 2.0) software framework (Blom et al., 2016).

4.9.3 Identification of conserved proteins of MRSA and subtractive genomics

Paralogs or redundant sequences from the *S. aureus* core-proteome were removed using CD-HIT (Li & Godzik, 2006) with a sequence identity cut-off of 0.8 (80%), and further core essential proteins of *S. aureus* were analyzed using GEPTOP 2.0 with essentiality score cut-off of 0.24 (Wen et al., 2019). The essential non-paralogous protein sequences were subjected to BLASTp against the genome of *Homo sapiens* (Gasteiger et al., 2003) using default parameters. The resultant sequences showing significant similarity with the Human host were discarded, while non-homologous sequences with no hit found were selected for subsequent analysis.

4.9.4 Characterization and prediction of subcellular location of proteins

The essential non-host homologous protein sequences were used for the prediction of drug targets and vaccine candidates. The non-host homologous protein sequences were used for the prediction of subcellular location using optimized PSORTb 3.0 (Yu et al., 2010) assembled in Vaxign 2 tool (He et al., 2010). These tools provide information on the subcellular localization of proteins as cytoplasmatic (CYT), secreted (SEC), potentially surface exposed (PSE), and membrane (MEM).

4.9.5 High-throughput structural modelling

The CYT protein sequences were submitted to MHOLline 2.0 server (Rossi et al., 2020) to model three-dimensional (3D) cytoplasmic proteins. This software integrates with HMMTOP, BLAST, BATS, MODELLER, and PROCHECK to analyze and classify potential drug targets based on their structural quality. Those proteins from the G2 group defined by BATS program having very high, high- and good-quality proteins were selected for the next stages of molecular docking.

4.9.6 Druggability analysis of drug targets

The selected drug target proteins from the G2 group were analyzed for druggability using DoGSiteScorer (Volkamer et al., 2012), an automated pocket detection and analysis tool for calculating the druggability of protein cavities. The druggability cavity of each drug target with a druggable score greater than 0.8 was selected for the docking analysis. The molecular weight (Mol. Wt.) of the target proteins was analyzed using ProtParam tool (Gasteiger et al., 2003). Eight best drug targets were chosen based on the above parameters and further analyzed for molecular function (MF) and biological process (BP) using UniProt (Consortium et al., 2021)

4.9.7 Ligand libraries, virtual screening, and molecular docking analysis

The ligand, drug-like molecules mentioned in the article (Vilela et al., 2019) were obtained in structural data format (.SDF) from the ZINC 15 database (https://zinc15.docking.org/) (Sterling & Irwin, 2015), and the *in-house* library was constructed. The 3D structures of all the target proteins were inspected for structural errors such as wrong bonds, missing atoms, and protonation states in Molegro Virtual Docker (MVD) 6.0 (Bitencourt-Ferreira & de Azevedo, 2019). The cavities generated by MVD for each target were compared with the cavities detected by DoGSiteScorer (druggability ≥ 0.80). The most druggable cavity defined by MVD was subjected to virtual screening. The MolDock Optimizer search algorithm was used in this analysis, which is based on a differential evolutionary algorithm, using the default parameters, which are (a) population size = 50; (b) scaling factor = 0.5, and (c) crossover rate = 0.9. The 3D poses and 2D representation of docked molecules were analyzed in PyMOL 2.4.1 (Schrödinger & DeLano, 2020) and Discovery Studio Visualizer 2020 (BIOVIA, 2020). Additionally, the Simplified Molecular Input Line Entry System (SMILES) format of drug molecules was analyzed for their ADME/pharmacokinetic

profile and drug-likeness parameters using the swissADME webserver (Daina et al., 2017).

4.9.8 Reverse vaccinology approach for prediction of putative vaccine candidates

The proteins localized in the extracellular or outer membrane regions of the organism are highly important for reverse vaccinology since these proteins are the first to be in contact with host immune cells and stimulate immune responses (Sanchez-Trincado et al., 2017). The non-CYT proteins were submitted to the Vaxign 2.0 tool (He et al., 2010) to predict the major histocompatibility complex (MHC I) and (MHC II) binding properties with adhesion probability greater than 0.5, the number of transmembrane helices, and no similarity to host proteins. B-cell epitopes were predicted using SVMTriP with epitope length 20 amino acid (aa) parameter (Yao et al., 2012). The antigenicity of vaccine candidate proteins was predicted using the VaxiJen v2.0 server with a threshold of 0.4 (Doytchinova & Flower, 2007). The instability index and molecular weight (Mol. Wt.) of the candidate proteins were analyzed using ProtParam tool (Gasteiger et al., 2003). The sequence-based domain information of vaccine candidates was retrieved from the conserved domain database, CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al., 2015).

5 Results and Discussion

5.1 Phenotypic and genotypic characterization

The colony characteristics of the strain are yellow colored, moist, round, glistening opaque colonies with β or weak hemolysis on blood agar. The strains are Gram-positive cocci showing a typical staphylococcal bunch. The screening of 35 *S. aureus* clinical strains revealed that 100% of the strains were resistant to cefoxitin. The ART result indicated that 94.28% of the strains were resistant to oxacillin. **(Table S1)**. The biochemical test, AST, and *S. aureus* species-specific sequence detection revealed concordant results **(Fig. S1)**. AST and detection of *mecA* gene results showed 33 (94.28%) strains were MRSA while 2 (5.72%) strains (SA G3 and SA G7) were *mecA* gene negative strains **(Figs. S2 & S3)**. The SA G3 and SA G7 strains showed the borderline-resistant to cefoxitin antibiotic. According to the CLSI 2014 M100 suggested that cefoxitin disk diffusion testing is more reliable for MRSA phenotype detection than oxacillin (CLSI 2014) due to its high sensitivity and specificity in

identifying *mecC* positive MRSA strains (Skov et al., 2014). It was reported that *mecA* PCR tests could not identify the strains possessing novel resistance mechanisms such as *mecC* gene or uncommon phenotypes (borderline-resistant oxacillin/ cefoxitin resistance) (https://www.cdc.gov/mrsa/lab/index.html). In this study, these two strains could possess the *mecC gene* (Fernando García-Garrote et al., 2014). Among the tested strains, 26 strains (71.43%) were found to be resistant to erythromycin. The prevalence of erythromycin-resistant strains collected from Hungary representing 90.47% is in good agreement with previous data (Cameron et al., 2016). It was also reported that the frequency of erythromycin resistance from 2010-to 2015 was 72% in Germany (Lina et al., 1999) and 60% in Greece (Holmes et al., 2005). In the present study, 24.24% (8/33) strains were found positive for the *pvl* gene **(Figs. S2 & S3)**. This gene plays a role in the pathogenicity of *S. aureus* by provoking necrosis, accelerating apoptosis, and destruction of polymorphonuclear and mononuclear cells, thereby contributing to morbidity and mortality (Kaya et al., 2018). It was reported that the low prevalence of *pvl* has been found at 5% and 4.9% in MRSA strains isolated from France and the UK respectively (Moroney et al., 2007). Among these strains, three carried *SCCmec* type IV and one classified as *SCCmec* type V, however, four strains carried *SCCmec* II (**Table S2**). According to previous reports, MRSA strains belonging to *SCCmec* types I, II, and III are dominant among the HA-MRSA, while *SCCmec* types IV and V are characteristic of CA-MRSA (Rossney et al., 2007). *SCCmec* typing revealed that *SCCmec* type II (33.33%) was the most predominant followed by type I (21.21%), type IVa (15.5%), type IVd (9.09%), IVb (3.03%), and the rest of the strains were non-typeable (12.12%). Among the 35 strains, eight strains were positive for *pvl* of which four strains belonged to *SCCmec* II, three strains to *SCCmec* IV, and one to *SCCmec* V **(Table S2 & Fig. S4)**. The distribution of *SCCmec* types showed that 36.36% of strains were CA-MRSA and 54.54% strains were HA-MRSA (**Table S1**). It was also reported that *SCCmec* type II usually presents in multidrug-resistant MRSA strains (Chua et al., 2014; Monecke et al., 2011) and were dominant outside European countries (Cheng et al., 2010). Our data related to *SCCmec* type IV showed a higher prevalence (27.6%). The reason behind this observation is probably due to the easy acquisition of short-size *SCCmec* type IV cassette (Goh et al., 1992; Schwarzkopf & Karch, 1994). Even though the representation of non-typeable *SCCmec* in our case complies with the previous finding (Martineau et al. 1996; Hookey et al. 1998; Shopsin et al. 2000; Mahmoudi et al. 2017), a few non-

typeable *SCCmec* can be reduced by applying the new *SCCmec* cassette detection (Mahmoudi et al., 2017). Some of the MRSA strains that harbored *SCCmec* IVa showed signs of *pvl* gene negative (**Table S2**), which is similar to the finding reported earlier (Janwithayanuchit et al., 2006). Our data related to *pvl* gene detection revealed that MRSA strains harboring *SCCmec* IVb and *SCCmec* IVd were found negative. Our finding also supports the idea that the harboring bacteriophage *pvl* gene by MRSA strains may not be a promising marker for CA-MRSA (Shakeri et al., 2010). This conclusion is supported by other studies about *SCCmec* typing for the classification of HA-MRSA and CA-MRSA (Cheng et al., 2010). Taken together, our findings suggested that the *SCCmec* typing method is more informative in problem-solving approaches (control and prevent infections caused by MRSA strains) for clinicians and epidemiologists.

The biofilm produced by *S. aureus* allowed the bacterial cell to resist immune system clearance and antimicrobial agents. The biofilm adheres to implanted biomedical devices and causes device-associated infections (Mack et al., 2004). Biofilm mass was increased at glucose concentration with a threshold response of 0.24% for *S. aureus* (Wardrop et al., 2014). Also, biofilm formation analysis of *S. aureus* using 0.25% glucose in medium showed an increase in the Cell Index (CI) signal for the biofilm-producing *S. aureus* than non-biofilm producing *S. aureus* by xCelligence real-time cell analyzer (Gutiérrez et al., 2016). Using glucose at 0.25% showed significantly enhance quorum sensing, biofilm formation, protease production, and swarming and swimming motility (Jahid et al., 2013). In consequence, the biofilm production was quantified using TSB supplemented with 0.25% glucose to enhance the biofilm production by *S. aureus* isolates. The quantitative test for biofilm production revealed that among the 33 *mecA* positive-MRSA strains, 87.87% of strains produced biofilm. Fifteen strains were selected among 33 MRSA strains based on non- (26.7%), moderate- (40%), and strong- (33.3%) biofilm-forming abilities for molecular typing and discrimination **(Table S4)**.

5.2 Molecular genotyping

S. aureus secretes the coagulase enzyme, a polypeptide that helps promote the clotting of plasma or blood (Asadollahi et al., 2018). The *coa* gene shows heterogenicity in the 81 base-pair long tandem repeats region differing in number and location of restriction sites among the *S. aureus* isolates (Omar et al., 2014). The assay

based on PCR amplification of the *coa* gene followed by RFLP was used to differentiate among the geographically diverse MRSA strains. This technique is simple, rapid, specific, inexpensive, and reproducible; allowing early recognition of an epidemic strain in a hospital setting (Koreen et al., 2004). In this study, *coa*-PCR typing yielded seven different amplicons in a size range from 550 bp to 800 bp. Among 15 *S. aureus* strains, the highest occurrence size is 700 bp (33.33%) as shown in **Table S6 & Fig. S5**, however, it was earlier reported that 600 bp amplicon is the most predominant (Strommenger et al. 2008). The result of the *coa* gene PCR–RFLP is summarized in **Table S6 & Fig. S6**. To get more insight into the similarity and difference of complex RFLP banding patterns, presence/absence heat-map and dendrogram were generated. Visualization of banding patterns revealed six distinct clusters, namely A-F with a calculated prevalence of 6.6, 13.3, 13.3, 20, 33.3, and 13.3%, respectively **(Fig. S7)**. Typing of *coa* gene and *HaeIII* RFLP, as well as DIs, were presented in Table S6. Discrimination of *coa* gene-specific amplicon pattern was further improved by *HaeIII* restriction enzyme digestion, which yielded 11 types of patterns with DI of 0.9619. Our data were in good agreement with the previous result in which DI was improved by digestion (Peacock et al. 1999; Rohde et al. 2007; Zmantar et al. 2008).

S. aureus produces protein A, an antiphagocytic protein that is coded by the *spa* gene (Ghasemian et al., 2015). Typing of the *spa* gene revealed 10 amplicons, ranging in size from 355 to 560 bp **(Fig. S8)**. The analysis of *spa*-sequence revealed high diversity, however, two strains belonged to *spa*-type t008, and another two were classified as *spa*-type t062 **(Table S7)**. In a previous study conducted in German, *S. aureus* isolates reported that t003 and t008 were predominant *spa* types (Fitzpatrick et al., 2005; Møretrø et al., 2003). Also, a recently published article stated that t008 was the most prevalent *spa* type in Europe and America (Galdbart et al., 2000).

The *spa*-PCR typing method produced eleven different genotypes with variable amplicons size ranges from 335 to 560 bp and revealed 0.9429 DI, which provides similar DI with the *coa*-*HaeIII* RFLP method which is supported by a previous study Peacock et al. (2000). Analysis of the *spa* gene revealed twelve known *spa* types. SA G11 & SA H29 strains possessed the *SCCmec*-IV gene (CA-MRSA), and both were in the same t008 *spa*-type. SA H16 & SA H19 strains clustered together into t062 *spa*-type, and both were found resistant to erythromycin, also classified into the HA-MRSA group **(Table S7)**. This study suggested that *spa*-typing has performed better

than other molecular typing methods and showed better DP. This typing method is useful for studying the genetic diversity of *S. aureus* for the epidemiological tracking of the source of infections (Feng et al., 2008; Gordon & Lowy, 2008; Tenover & Gaynes, 2000) and offers several advantages in comparison with alternatives methods, such as a publicly available comprehensive and curated database for analyzing *spa* sequence with standard nomenclature (Pope, 2019).

5.3 Molecular identification of biofilm-related genes

Although, all the selected strains did not produce biofilm each of them harbors genes for intracellular adhesion (*icaADBC*) and regulation (*icaR*). The presence of *fnbA* and *fnbB* genes were detected in 73.3% and 66.6% strains, respectively. Genes associated with biofilm-forming ability *viz*., *cna*, *clfA*, *clfB*, and *ebps* were found present in 53.33%, 80%, 73.3%, and 86.6% strains, respectively **(Figs. S9–S19)**. Earlier, the prevalence of the *cna* gene was reported to range from 22% to 56.5% (De Paepe et al., 2014). The *clfA, clfB,* and *epbs* genes play an initial role in biofilm development (Guttman et al., 2005), however, our data showed that the presence or absence of these genes does not represent a clear discriminative marker for differentiating strains in terms of biofilm-forming ability. In good agreement with previous data, we found that not all *ica*-positive isolates produce biofilms (Clokie et al., 2011). In this study, we observed that 66% of the isolates harbor two *fnb* genes almost similar to the results reported by (Chambers & DeLeo, 2009) for European *S. aureus* strains. However, the presence of *fnb* genes in an isolate does not guarantee the biofilm-forming ability of the isolate. Thus, the presence or absence of biofilm-related genes does not represent a clear discriminative marker for differentiating strains in terms of biofilm-forming ability.

5.4 Cluster analysis based on a polyphasic approach

Although, the individual test, for example, performed AST, catalase, coagulase, DNase, citrate utilization, urease production, mannitol fermentation, blood lysis, and biofilm production assays has the advantage of being cost-effective, but often cannot differentiate among the strains. PCR-based detection of *mecA* gene and genes responsible for PIA (*icaADBC* and *icaR*) could not differentiate the clinical isolates in the present study. PCR-based detection of *pvl* gene and genes encoding for MSCRAMMs (*fnaA, fnaB, clfA, clfB, cna*, and *ebps*) also showed poor DP, while *SCCmec, coa-Hae*III RFLP, and *spa* typing revealed moderate DP.

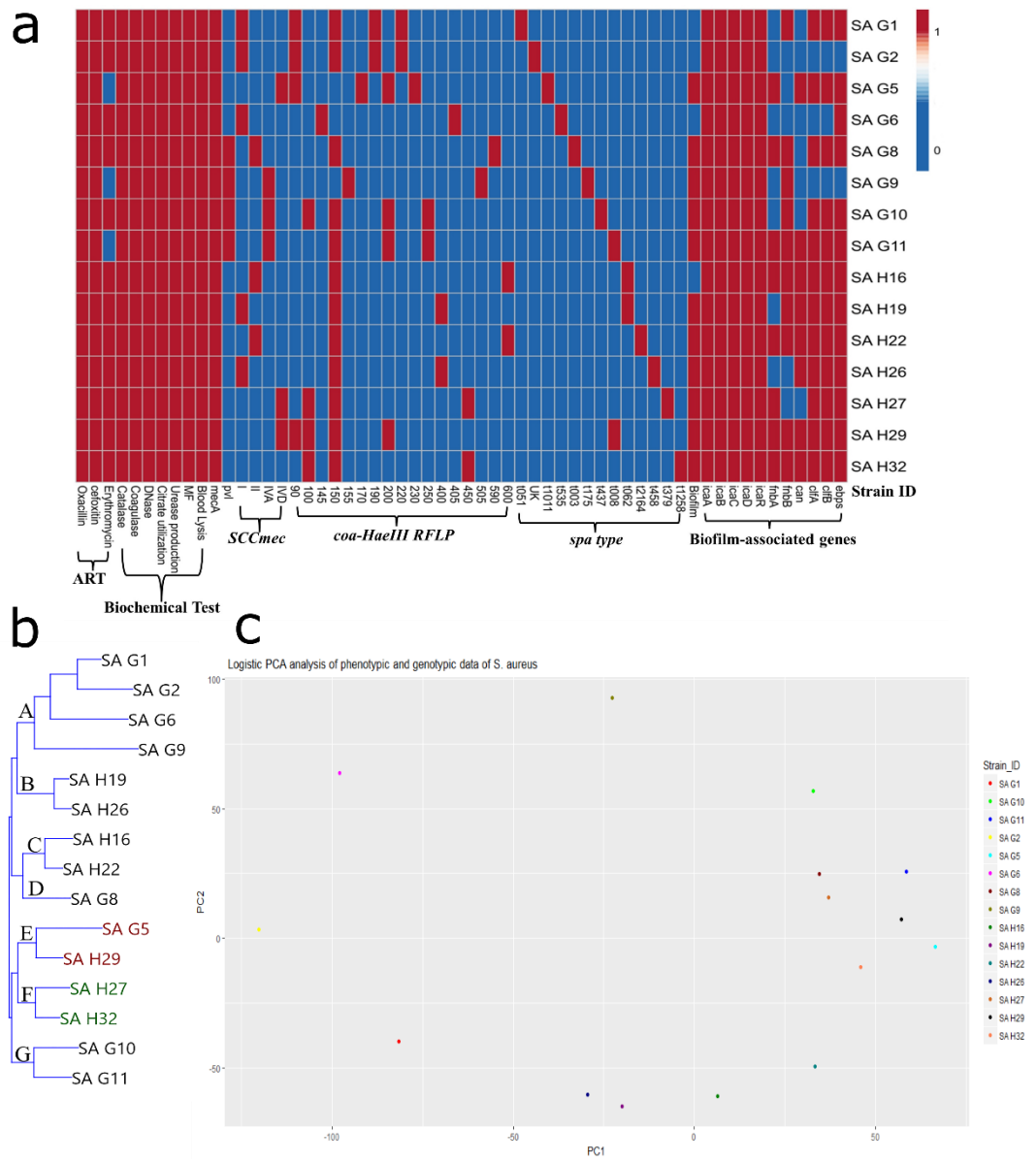


Fig. 2. Polyphasic comparison of *S. aureus* clinical isolates. (a) Colored similarity matrix category representing resistant test against different antibiotics (ART), biochemical test, PCR based *mecA* and *pvl* genes detection, *SCCmec* typing, *coa-HaeII*I RFLP, *spa* type, biofilm-forming assay, and detection of biofilm-associated genes. Red and blue colors indicate the presence (1) and absence (0) of particular properties, respectively. **(b)** NJ- dendrogram prepared from the results of biochemical and molecular analysis of the 15 *S. aureus* clinical isolates (SA). The isolates were labeled according to their geographical origin where G and H indicate Germany and Hungary. Clusters are labeled as A-G. **(c)** Principal component analysis of the results of biochemical and molecular analysis of the 15 *S. aureus* clinical isolates.

To avoid a misleading conclusion, the data from all applied methods were coupled to perform cluster and PCA analysis. The generated dendrogram and PCA analysis suggested that Hungarian strains (SA H27 & SA H32) belonged to the same *SCCmec*-IV and shared the same cluster F, however, these strains were isolated from the different sites of infections (nostrils and trachea) and showed different antibiotic resistance patterns and biofilm-forming abilities. Except for cluster E, Germany and Hungarian strains (SA G5 & SA H29) from different geographical locations clustered together. However, the strains collected from Germany *viz*., SA G6 & SA G8 belonged to different *SCCmec* types and had similar antibiotic resistance patterns and biofilm-forming profiles, but these strains were isolated from the different sites of infections (skin and other body sites) and not clustered in the same group **(Figs. 2b & 2c)**.

5.5 General genomic features of *S. aureus* isolates

The polyphasic characterization data showed that the strains originated from the same geographical region were found in the close group (SA G6 and SA G8; SA H27 and SA H32) while the other two strains originated from different geographical regions *viz*., SA G5 (German strain) and SA H29 (Hungarian strain) were found in the same group **(Fig. 2a–c)**. Based on this information, these six *S. aureus* strains were chosen for in-depth comparative genome levels study to better understand the genomic differences among the strains.

The genomic DNA of *S. aureus* isolates was successfully sequenced in the IonTorrent PGM sequencing platform. The average raw reads obtained from the genome sequencing of six *S. aureus* genomes are ~ 57.8, 88.9, 69.6, 48.9, 128.3, and 92.7 million bases (Mb) for genomes of SAG5, SA G6, SA G8, SA H27, SA H29, and SA H32 strains, respectively. The generated reads per sample cover more than 98% of the reference genome (ASM1342v1) with an average depth of 152.4X. The closely related strains identified by kmerFinder 2.0 were *S. aureus* CA-347 (CP006044.1), *S. aureus* subsp. *aureus* ST228 (HE579073), *S. aureus* subsp. *aureus* JH9 (CP000703), *S. aureus* subsp. *aureus str.* Newman (AP009351.1) for SA G5, SA G6, SA G8, and SA H29 strains, respectively. Also, *S. aureus* subsp. *aureus* HO 5096 0412 (HE681097.1) was identified as closely related strains for SA H27 and SA H32 strains. Among the *S. aureus* isolates, SA G8 has the largest genome size (28633393 bp) with a high % GC content (32.81%). The numbers of protein-coding sequences (CDSs) in the *S. aureus* strains varied from 2630 (SA H27) to 2743 (SA H29).

Table 1. General genomic features of *S. aureus* genomes in this study

Strains	**SA G5**	**SA G6**	**SA G8**	**SA H27**	**SA H29**	**SA H32**
Size (bp)	2760385	2856214	2857863	2783185	2834624	2786627
Contigs	36	103	83	44	47	63
Scaffolds	1	22	15	1	3	3
N50 (bp)	369923	125160	263953	328241	243826	208577
GC%	32.77	32.79	32.81	32.73	32.65	32.72
CDS	2689	2734	2743	2630	2843	2657
Genes assigned to SEED	1993	2101	2080	2014	2159	2036
rRNA	9	9	10	8	6	9
tRNA	59	61	60	57	51	60
Prophage Regions	3	3	5	3	3	1
[#]Plasmids	-	p1G6	-	-	p1H29, p2H29	p2H32
SCCmec type	IVd	I	II	IVa	Vb	IVa
MLST	ST45	ST228	ST225	ST22	ST8	ST22
Spa type	t1001	t535	t003	t379	t008	t1258
agr-type	I	II	II	I	I	I
Accession no.	CP032160	RAHA00000000	QZFC00000000	CP032161	CP032468-CP032470	RAHP00000000

[#]**Plasmids:** The presence of plasmid in a genome is indicated by plasmid name, while absent is represented by a minus (-) sign.

The comparison of draft genome assemblies, genome annotation, molecular typing, plasmid, and prophage features for *S. aureus* genomes was summarized in **Table 1**. The molecular epidemiology study of MRSA helps to find the risk factors associated with MRSA infections and is able to differentiate the several MRSA strains (Monecke et al., 2011). The genome-based molecular epidemiology studies found that German isolates exhibit *SCCmec* type IVd with ST45, *SCCmec* type I with ST228, and *SCCmec* type II with ST225 while Hungarian isolates hold *SCCmec* type Vd with ST8, and *SCCmec* type IVa with ST22. Also, *agr* type II and I were observed in Germany and Hungarian isolates, respectively **(Table 1)**. Previous studies suggested that MRSA strains with *SCCmec* type I or II or III are dominant among the HA-MRSA, while *SCCmec* types IV or V are the characteristic of CA-MRSA (Bhutia et al., 2015; Shukla et al., 2012). The STs of Hungarian isolates belonged to CC8, and CC22 suggesting its relationship with CA-MRSA, while Germany isolates SA G6, and SA G8 belonged to CC5 which is typical of HA-MRSA. In hospitals, the multidrug-resistance *SCCmec* type III was replaced by the multidrug-susceptible *SCCmec* type IV (ST22) strains slowly (D'Souza et al., 2010). SA G8 isolate was found positive for the *pvl* gene, which is commonly used as a marker of CA-MRSA (Barrera-Rivas et al., 2017; Lina et al., 1999) besides this toxin has been shown to play a role in necrosis, accelerating apoptosis, and polynuclear- and mononuclear cells lysis, thereby contributing morbidity and mortality (McCarthy & Lindsay, 2012).

5.6 Genes encoding plasmids

The putative plasmids were identified in nonaligned contigs that displayed an unexpected high coverage level after the genome assemblies. A putative plasmid (p1G6) of 13,331 bp length was identified at Scaffold 4 of the SA G6 genome consisting of the replication gene (*repA*). The p1G6 plasmid has 30.97% sequence coverage with plasmids pTW20_1 (FN433597.1) **(Fig. 3a)**. The sequence coverage region of p1G6 with pTW20_1 constitutes the genes that encode for proteins such as IS6 family transposase, replication-associated protein (Rep), cadmium resistance transporter (CadD), cadmium efflux system accessory protein (CadX), replication initiation protein A (RepA), quaternary ammonium compound efflux MFS transporter (QacA), multidrug-binding transcriptional regulator (QacR), DUF536 domain-

containing protein (mP), AAA family ATPase (Abp), hypothetical proteins, HAD hydrolase family protein, and IS257 family transposase.

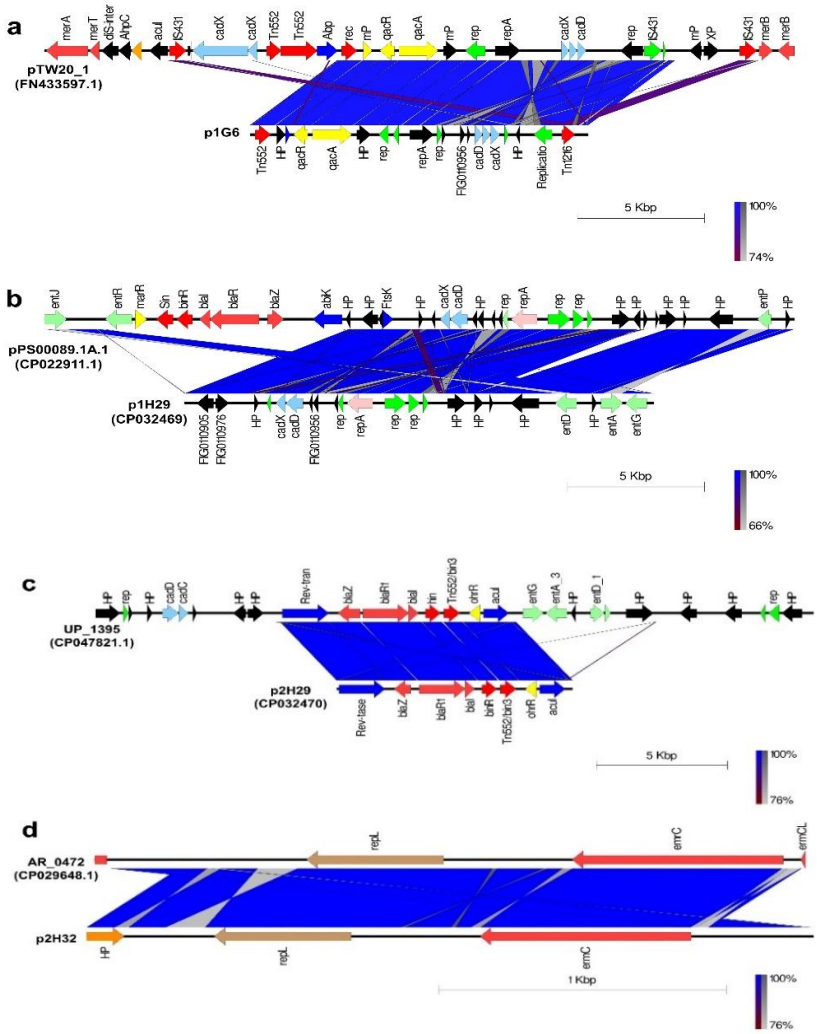


Fig. 3. Comparison of linear plasmid maps by Easyfig alignment. Coding Sequences are represented by colored arrows. Blue lines between the plasmids indicate the shared similarity regions according to BLASTn identity. CDS are characterized by functions as follows: antiseptic resistance genes (yellow), erythromycin resistance gene (light red), DNA replication (green), transposons/integrases (red), replication A gene/hypothetical proteins/others (black), replication L gene (brown), and cadmium resistance gene (cyan). The outer scale is marked in kilobases. (a) Sequence alignment of p1G6 plasmid with the reference pTW20_1 plasmid, (b) p1H29 plasmid with the reference pPS00089.1A.1 plasmid, (c) p1H29 plasmid with the reference UP_1395 plasmid, and (d) p2H32 plasmid with the reference AR_0472 plasmid.

Two putative plasmids, p1H29 and p2H29 of 17,165 bp and 9020 bp lengths, respectively were identified in nonaligned contigs (scaffolds 2 and 3) of SA H29 genome, and these two plasmids constitute *rep20* and *rep7C* type genes **(Table 1)**. Plasmid p1H29 has showed 44.20% and 57.34% sequence coverage with plasmids pBU108b (KF831356.1), and pPS00089.1A.1 (NZ_CP022911.1), respectively **(Fig. 3a)**. Plasmid p2H29 has a length of and showed 23.42% and 23.45% sequence coverage with plasmid: II (LT671860.1), and UP_1395 plasmid (NZ_CP047821.1) respectively **(Fig. 3b)**. Identified plasmid p1H29 carried genes encoding cadmium resistance (CadD) and transportation (CadX) proteins, and enterotoxins (EntA, EntD and EntG). Identified plasmid, p2H29 has *blaZ* (beta-lactamase), *blaR*, and *mecI* genes that conferred resistance to penicillin **(Fig. 3c)**. The SA H32 genome also consists of a putative plasmid (p2H32) having a length of 2530 bp located at Scaffold 3 and showed 71.32% sequence coverage with plasmids AR_0472 (NZ_CP029648.1). It consists of a replication gene (*repL*) and carried an erythromycin resistance gene (*emrC*) **(Fig. 3d)**. The identified plasmids of *S. aureus* encode no other factors for their transfer, such plasmids may transfer *via* phage generalized transduction (LaBreck et al., 2018). The linear graphical map of plasmid comparison was represented in **Fig. 3**. The plasmid p1G6 carried *qacA* gene, which is known to decrease chlorhexidine (antiseptic) susceptibility and give an event of MGEs transfer evidence of *qacA* across the *S. aureus* strains (McCarthy et al., 2014; McCarthy & Lindsay, 2012). The harbor of MGEs (mosaic features of prophages and plasmids) contributes to the tremendous distribution of ARGs and VFGs among the *S. aureus* isolates (Lindsay, 2010). This MGEs transfer event could be useful for the survival of *S. aureus* in different ecological niches (van Wamel et al., 2006).

5.7 Characteristic of prophages-like elements

The genomes of *S. aureus* isolates have several prophages and phage-like element regions, and these prophages belonged to the *Siphoviridae* family and had temperate lifestyles. The highest number of prophage regions was found in the genome of SA G8 isolate including three intact prophages, a questionable, and an incomplete prophage. Four prophage regions were found in the genome of SA G6 isolate including an intact prophage, two questionable prophages, and an incomplete prophage. The genome of SA H27 isolates harbor three intact prophages while the genome of SA H32 harbor only one intact prophage. The *lukF-PV* and *lukM* genes

(Bicomponent leukotoxins), and *plc* gene (phospholipase C) were identified in the prophages of SA G6, SA G8, SA H27, and SA H32. The family of beta-hemolysin converting phage encodes proteins such as SCIN (staphylococcal complement inhibitor) and CHIPS (chemotaxis inhibiting protein of staphylococcus) involved in host-pathogen interaction and contribute to evading human innate immune response (Barrera-Rivas et al., 2017). The prophages of SA G6, SA G8, and SA H27 carried *sak* gene (staphylokinase) and *scn* gene (staphylococcal complement inhibitor). Chemotaxis inhibitory protein encoded by *chp* gene was identified in SA G8 and SA H27 prophages. Therefore, prophages were the reservoir of virulence and resistance factors that play a role in the evolution of virulence strains and cause a major threat to human and animal health (Barrera-Rivas et al., 2017).

Enterotoxin A encoded by *sea* gene was harbored by the prophages of SA G6 SA H27. Hemolysin genes such as *hlb* (β-hemolysin), and *hlgB* (γ-hemolysin B) were found in the prophages of SA H27, and SA H32. In addition to virulence factors, phiG6.4 prophage carried ARGs genes that conferred resistance to beta-lactamase (*blaZ*), aminoglycoside (*ant(6)-Ia*, and *aph(3*0*)-III*), and nucleoside (*sat-4*) antibiotics. Also, it was detected the presence of *blaZ* gene in SA G8 **(Table S8)**. It was reported that the excessive administration of antibiotics in the hospital can result in prophage induction due to antibiotic-induced SOS response. Under selective pressure exerted by antibiotics, the prophages are excised along with the bacterial DNA (ARGs) and packed into a capsid. These ARGs are incorporated into phages and will continue their mobilization by HGT through the animal biomes and human biomes. In some bacterial strains, prophages are considered as drivers of gene transfer especially ARGs (Colavecchio et al., 2017). This allows rapid exchange of ARGs in *S. aureus* particularly and in other pathogenic bacteria and may help promote the emergence of MDR strains which ultimately leads to widespread prevalence. The presence of ARGs and VFGs in the prophage regions of SA G6 genome differentiates it from the other *S. aureus* isolates and may determine its greater pathogenic potential by modifying its antigenicity (Maddux, 1991). The comparative analysis of VFGs associated with putative prophages was summarized in **Table S8.**

5.8 In-silico analysis of antimicrobial resistance and associated genes in the genomes

Six study genomes of *S. aureus* strains shared 63.33% (19/30) antibiotic resistance and associated genes **(Fig. 4)**. The shared genes consist of methicillin-

resistant PBP2a (*mecA* and *mecR1*); multidrug resistance efflux (*ygaD*); fluoroquinolone (*norA*); fluoroquinolone and acridine dye (*arlS* and *arlR*); glycylcycline (*mepA*); tetracycline (*tet-38*); tetracycline, penam, cephalosporin, glycylcycline, rifamycin, phenicol, triclosan, fluoroquinolone (*mgrA* and *marR*); lipopeptide (*pgsA*, *clsA* and *rpoC*); rifampicin (*rpoB*); aminocoumarin (*gyrB* and *parE*); dihydrofolate reductase (*dfrA*/*folA*) and defensin (*mprF/fmtC*, multiple peptide resistance factor) that play roles in resistance mechanism including antibiotic efflux, antibiotic target alteration, and antibiotic target replacement. Further, *gyrA*, *grlA*, and *grlB* genes mutation were detected in the six study genomes and reference genomes. The mutation in these genes is associated with quinolone resistance (Schmitz et al., 1998). The closest reference genomes predicted by kmerFinder 2.0 were compared with the study genomes to explore the similarity among them in terms of ARGs acquisition. This study showed that the genome of SA G6 isolate acquired the highest numbers (26) of ARGs, while *S. aureus* subsp. *aureus str.* Newman (AP009351.1) acquired low numbers (17) of ARGs (**Fig. 4**). Staphylococcal β-lactamase encoded by *blaZ* gene is carried by the transposon Tn*552* or Tn*552*-like elements located on a large plasmid and can be non-inducible or inducible with antibiotics (Pugazhendhi et al., 2020). It was noticed that *blaZ* gene was absent in SA G8 isolate, probably due to the curing of *blaZ* positive plasmid (Lim et al., 2014). Erythromycin resistance gene (*ermA*) was detected in the chromosome of SA G6, SA G8, and SA H29 isolates, however, *emrC* gene was found in the plasmid of SA H32 **(Fig. 3d)**. It was suggested that these genes may not be involved in the loss of specific ARGs for environmental adaptation, but it is expected to be essential for these isolates (Bartlett & Hulten, 2010; Otto, 2014). This in-silico identification and antibiotic susceptibility test results were correlated with beta-lactam, and erythromycin antibiotic resistance analysis.

The other genes responsible for copper hemolysis, cobalt-zinc-cadmium resistant, arsenic resistance, mercury reductase, and mercury resistance operon were identified in the genomes of *S. aureus* isolates. The cadmium resistance was found in the SA G6 and SA H29 isolates. The secondary metabolite biosynthetic gene clusters identified among the genomes were staphylobactin, aureusimine, bacteriocin, and staphyloferrin A. The auto-inducing peptide-II gene was identified in SA G6 and SA G8 genomes while the auto-inducing peptide-I gene was identified in SA G5, SA H27, SA H29, and SA H32 genomes. However, lanthipeptide A (gallidermin) was present in SA H29 genome. The clinically isolated SA G6 carried an array of antibiotic and

heavy-metal resistance genes, this finding suggests that SA G6 could be a multi-drug resistant strain.

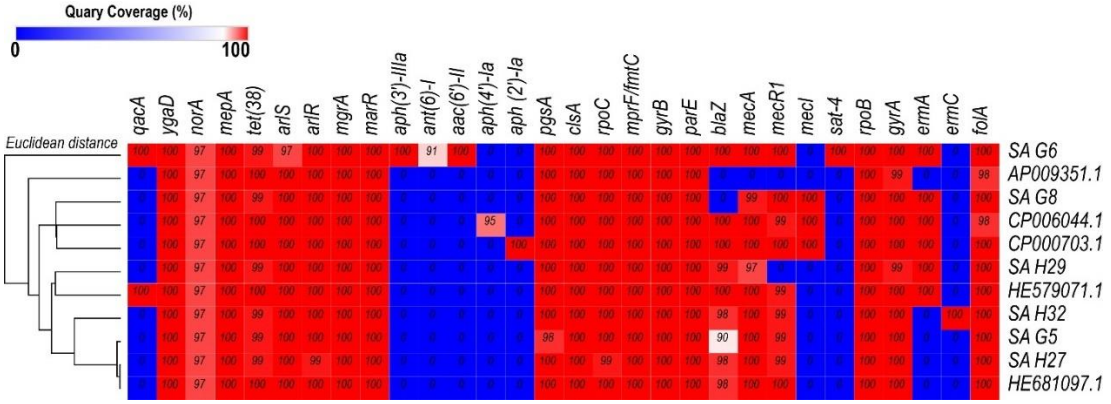


Fig. 4. Heat map showing the presence (red color) and absence (blue color) of antibiotic resistance genes. The labels on top indicate the gene names and the label on the left indicates the strains.

5.9 In-silico analysis of virulence-factors encoding genes in the genomes

The virulence-factors encoding genes were categorized into adherence, toxins, enzymes, ESAT-6-like proteins, heme-uptake, and immune invasion by VFDB. The VFGs of *S. aureus* genomes against the VFDB revealed that 40.98% (50/122) of genes were shared by all the strains of *S. aureus*. In the six study genomes of *S. aureus*, the shared virulence genes comprised 59.09% (14/22) adherence, 15.68% (8/51) toxins, 70.58% (12/17) enzymes, 33.3% (4/12) type VII secretary system, and 60% (12/20) immune invasion. The genome of SA G8 isolate has occupied 3.40% of VFGs against its CDS, whereas the genome of SA H32 isolate has 2.97% of VFGs against its CDS. Besides, among the compared strains, *S. aureus* subsp. *aureus str*. Newman (AP009351.1) has 3.59% (97/2701) of VFGs against its CDS.

The genomes of all isolates shared 15.68% (8/51) of toxin genes such as α-hemolysin gene (*hla*), β-hemolysin gene (*hlb*), δ-hemolysin gene (*hld*), γ-hemolysins (*hlgA*, *hlgB, hlgC*), exfoliative toxin type A (*eta*), and Panton-Valentine leukocidin (*lukF-PV*) **(Fig. 5a)**. The presence of *hlb* gene in the isolates contributes to the phagosomal escape of *S. aureus* and influences biofilm development (Vandenesch et al., 2003). The PVL toxin (*lukF-PV* and *lukS-PV*) was found in the prophage of SA G8 genome. This toxin has cytolytic activity against blood cells and leukocytes, contributing to the *S. aureus* pathogenicity (Argudín et al., 2010; Grumann et al., 2014). Staphylococcal enterotoxins (SEs) or staphylococcal superantigens proteins

(SAgs) are well-known for causing food poisoning, localized epidermal infections (bullous impetigo), and generalized diseases (Staphylococcal scalded skin syndrome) (Chen et al., 2004; Jarraud et al., 2001). SEs encoding genes are located on mobile elements including bacteriophages, pathogenicity islands (SaPI), or plasmids. In this study, the highest number of toxin genes were identified in the genome of SA G5 *i.e.,* 52. 94% (27/51), and in addition to shared genes, the extra genes were enterotoxin A (*sea*), enterotoxin B (*seg*), enterotoxin C (*sec*), enterotoxin I (*sei*) enterotoxin-like toxins (*selk, sell,selm, seln, selo, selu*), exotoxin (*set1, set2, set3, set4, set5, set7, set15, set16*), and Panton-Valentine leukocidin (*lukS-PV*). Cytolysin (*cylR2*) gene was acquired in the genome of SA G6 isolate. These *seg* and *sei* genes belong to *egc* (enterotoxin gene cluster), involve in staphylococcal food poisoning TSS, and SSF (Jarraud et al., 2001), and *egc* was distributed widely in clinical isolates and play a role in pathogenesis (Melish & Glasgow, 1970). In this study, *eta* gene which was shown to encode for ETA toxin was found in all the isolates. It is responsible for causing human skin damage and is most prevalent in Europe (Niemann et al., 2004; Pizarro-Cerdá & Cossart, 2006).

The most critical step in the infection process is the adhesion of the bacterial cell to the host cell surface where the adhesion molecules support the major role of bacterial cell survival and promote pathogenicity (Reffuveille et al., 2017). Adherence associated genes shared in six *S. aureus* isolates were 59.09 % (14/22) such as autolysin (*atl*), cell wall-associated fibronectin-binding protein (*ebh*), elastin binding protein (*ebp*), fibronectin-binding proteins A (*fnbA*), intercellular adhesin (*icaA*, *icaB*, *icaC*, *icaD*, and *icaR*), protein A (*spa*), and ser-Asp rich fibrinogen-binding proteins (*sdrC*, *sdrD*, and *sdrE*) **(Fig. 5b)**. Among the genome of isolates SA G8, and SA H28 present 77.27% (17/22) of adherence-associated genes with additional genes of clumping factor A (*clfA*), clumping factor B (*clfB*), and fibronectin-binding proteins (*fnbB*). MRSA is responsible for causing biofilm infections that are more difficult to treat and need more intensive care as compared to *Staphylococcus epidermidis* biofilm (Cramton et al., 1999). The principal component of biofilm formation is PIA which consists of different intracellular adhesion (*ica*) genes (Farran et al., 2013) and plays a crucial role in the initial stage of bacterial cell adherence to surfaces and intercellular adhesion for the cells to aggregate (Eckhart et al., 2007). These genes were detected in all isolates. However, the biofilm production ability ranging from weak to strong was observed. SA G6 isolate obtained from skin infection showed a very weak

biofilm-forming ability (**Table S4**). The low biofilm formation in SA G6 might be due to DNase enzyme found in skin cells (Nasr et al., 2012). The previous study revealed that the presence of the *ica* genes did not always correlate with biofilm (Cho et al., 2002; Kiem et al., 2004). Some authors reported that despite the presence of *ica* operon, some staphylococcal isolates produce weak biofilm production due to the inactivation of *icaA* by insertion of IS256 (Kleinert et al., 2016). A recent study reported that *sdrC* mutant exhibited significantly inhibited biofilm formation (Shin et al., 2013) and the expression of the *ica* operon and *sdrC* are responsible for biofilm formation (Boles & Horswill, 2008). Our study revealed the sequence variation in *sdrC* in Hungarian isolates, this variation might influence the biofilm formation. The global regulatory gene, *agr* repression has been associated with biofilm formation and its induction through AIP results in seeding dispersal in mature biofilm (Aires-de-Sousa, 2017). CA-MRSA strains showed higher activity of *agr*, which controls and enhances virulence (Zhang et al., 2018). It was reported that *S. aureus* strains that belonged to *agr* I group exhibited a stronger biofilm-forming ability than the strains that belonged to *agr* IV group (Huseby et al., 2010; Sugimoto et al., 2013). A similar result was observed in the current study isolate, SA H27. In addition to this extracellular adherence protein (encoded by *eap* gene), and beta toxin (encoded by *hlb* gene) play a role in biofilm maturation (Archer et al., 2011). Our data show that *eap* gene is present only in the SA H27 isolate. This gene might be responsible for high biofilm formation. Since biofilm formation involves PIA-dependent/independent pathways, regulator genes, and eDNA (Ghasemian et al., 2015). Also, the presence of such pathways in *S. aureus* may not provide much impact on biofilm formation profiling. There was a difference in the prevalence of biofilm-associated genes between the isolated strains. This suggests that the presence of genes encoding biofilm formation is not an absolute determinant in the ability of *S. aureus* to form a biofilm (**Fig. 2a**). Thus, our future studies will focus on the expression profiling of such relevant genes which may be necessary to determine the key genes involved in biofilm formation.

Several exoenzymes encoding genes namely cysteine protease/ staphopain (*sspB, sspC*), hyaluronate lyase (*hysA*), lipase (*geh, lip*) serine V8 protease (*sspa*), staphylocoagulase (*coa*), staphylokinase (*sak*), and thermonuclease (*nuc*) were present in the genomes of all isolates.

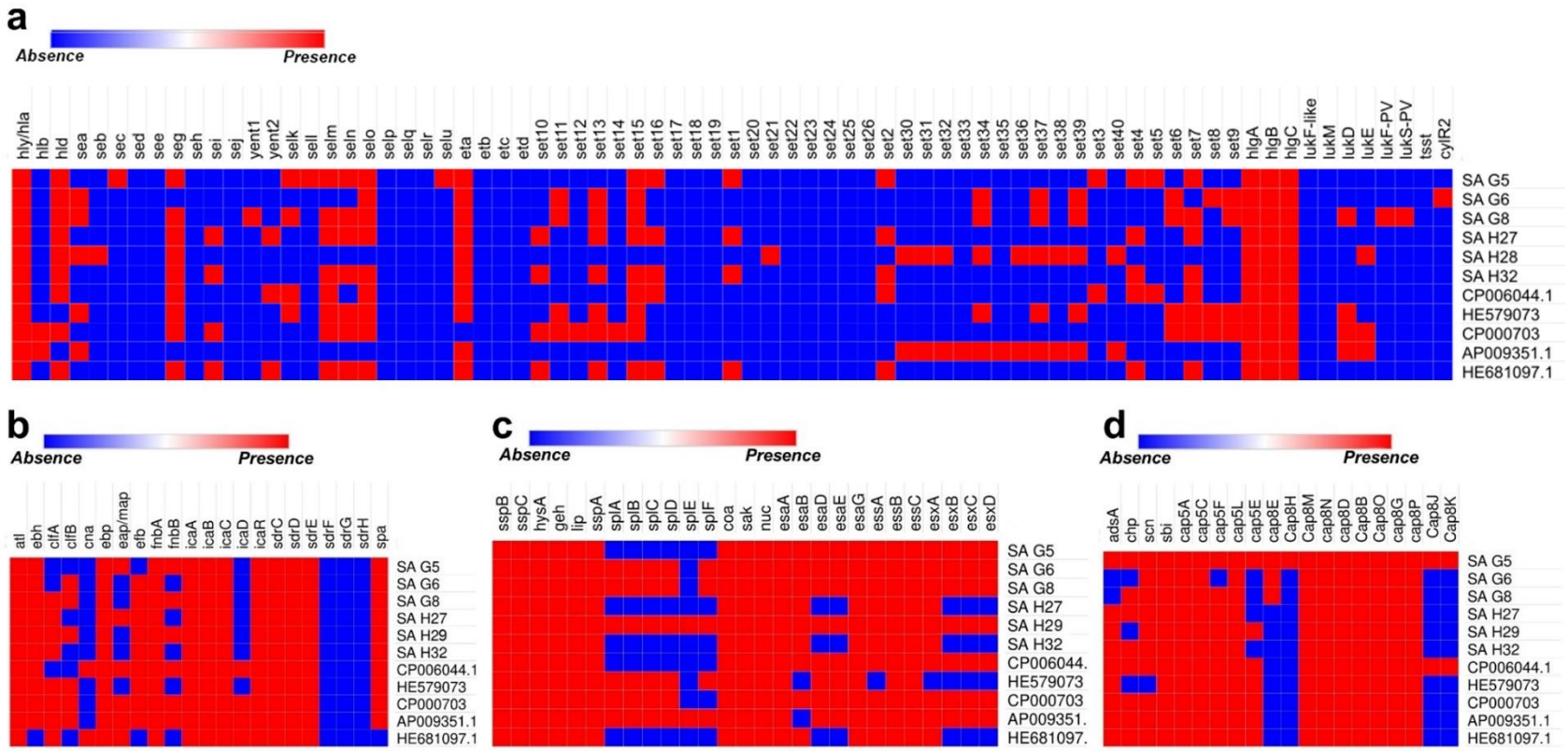


Fig. 5. Heat map showing the presence (red color) and absence (blue color) of (a) toxins; (b) adherence; (c) exoenzymes and secretary factors; and (d) anti-phagocytosis (capsules) factors. The labels on top indicate the gene names and the label on the left indicates the strains.

However, five genes cluster for serine protease (*splA, splB, splC, splD, splF*) were absent in the genomes of SA G5, SA H27, and SA H32 isolates **(Fig. 5c).**

Type VII secretary system consists of ESAT-6-like proteins which play role in the survival and ability to persist in their hosts (Burts et al., 2005). Type VII secretion system includes membrane-associated proteins (*esaA, essA, essB, and essC*), soluble cytosolic (*esaB, esaE, esaG*), and secreted virulence factors (*esxA, esxB, esxC, esxD,* and *esaD*) were identified in the genomes of Germany isolates and SA H28 isolates while *esaD, esaE, esxB, esxC,* and *esxD* genes were absent in the Hungarian isolates (SA H27 & SA H32) **(Fig. 5c)**. The *esxA* and *esxB* genes were shown to play a significant role in the distribution and colonization of *S. aureus*, as well as activation of the cell-mediated immune responses which result in increased pathogenesis (Cao et al., 2016). Also, *esaD* gene was found in German isolates and the SA H29 isolate. This suggests that it could inhibit the growth of other closely related *S. aureus* strains and play a role in an intra-species competition (Skaar et al., 2004). The iron-regulated surface determinants (*isd*) gene clusters involved in the iron uptake mechanism were identified in all the genomes of isolates. This system is used by *S. aureus* for stealing the iron from the heme group of hemoglobin and leads to hemolysis (O'Riordan & Lee, 2004).

Capsular polysaccharide synthesis genes namely capsular polysaccharide synthesis genes belonging to stereotypes 5 and 8 are predominantly detected in clinical isolates of *S. aureus* showing significant virulence. They target the antibodies that protect against Staphylococcal infections (Bosi et al., 2016; Sharma et al., 2018). In this study, all the genomes of isolates have capsular polysaccharide synthesis enzymes **(Fig. 5d)**. Other genes responsible for the host immune evasion such as IgG-binding protein (*sbi*), and staphylococcal complement inhibitor (*scn*), were identified in all isolates. The genome of isolate SA G5 carried an additional number of genes including capsular polysaccharide synthesis enzymes such as Cap5F (*cap8F*), Cap5L, Cap5E, Cap8M, Cap8J, Cap8K, adenosine synthase A (*adsA*), and chemotaxis inhibiting protein (*chp*).

5.10 Comparative genome analysis

The genome comparative analysis based on core genome ANI mean matrices result indicated that all the *S. aureus* strains showed >96.69% sequence identities. Among the core genomes of *S. aureus* isolates, the SA H27 genome showed the

highest sequence identities (97.3 to 99.9%), however SA G5 genome has 96.7 to 97.2% sequence identities to genomes of *S. aureus* isolates **(Fig. S20a)**. Further, the core genome AAI mean showed that HO 5096 0412 (HE681097.1) strain has maximum proximity to all the *S. aureus* strains whereas MRSA252 (BX571856.1) strain showed the distinct from other strains **(Fig. S20b)**. Pairwise genome comparison using ANIb showed that SA H27 and SA H32 genomes have 97.4% similarity in their sequence alignment and found a closer similarity to the genomes of HO 5096 0412 (HE681097.1) and H-EMRSA-15 (CP007659.1) strains.

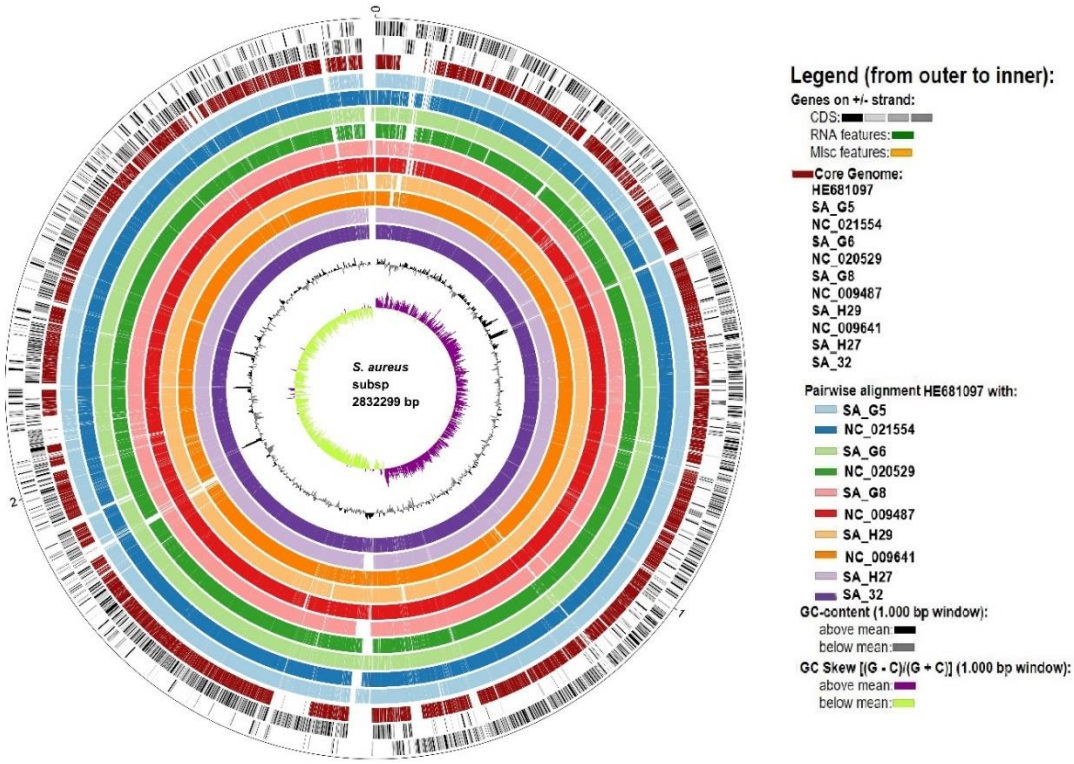


Fig. 6. Circular genome comparison map showing homologous chromosome segment of 10 *S. aureus* genomes with the reference genome of *S. aureus* subsp. *aureus* HO 50960412 (HE681097.1) strain. White spaces indicate regions with no identity to the reference genome, *S. aureus* subsp. *aureus* HO 50960412.

A whole-genome circular comparative map of 6 *S. aureus* genomes and their close reference genomes was generated against *S. aureus* subsp. *aureus* HO 5096 0412 (HE681097.1) genome using Biocircos based on BLAST sequence similarities. Each genome was indicated by a different color, and the darker areas in the circular genome showed a 100% sequence similarity with the reference genome, while the lighter (gray) areas showed a 70% sequence similarity **(Fig. 6).** The map revealed less gap

between the SAH27 (CP032161) and SA H32 (RAHP00000000) genomes showing high proximity between them when compared to other genomes. SA G5 (CP032160) genome has many gaps with white color than the other genomes showing a distant relationship.

5.10.1 Orthologue gene analysis

The SEED subsystem categories identified by RAST revealed that six study genomes of *S. aureus* possessed "amino acids and derivatives" (~16%) was the largest subsystem, followed by "carbohydrates" (12%), "protein metabolism" (10.5%), and "cofactors, vitamins, prosthetic groups, pigments" (~8.8%) **(Fig. 7a)**. The "carbohydrate", and "protein metabolism" subsystems were found largest in SA H29 (12.76%) and SA G5 (10.5%) genomes, respectively. Also, the SA G5 genome has the largest subsystem of "amino acids and derivatives" (16.01%). The subsystem belongs to "phages, prophages, pathogenicity island" (2.5%) was identified as the highest in the SA G8 genome. The SA G6 genome occupied the largest subsystem of "virulence, disease, and defense" (4.42%) such as adhesion, bacitracin stress response, colicin v, and a bacteriocin production cluster, copper homeostasis, bile hydrolysis, cobalt-zinc-cadmium resistances, multidrug resistances, 2-protein, mercuric reductase, mercury resistance operon, streptothricin resistance, teicoplanin-resistances, aminoglycoside adenylyltransferases, fluoroquinolone resistances, arsenic resistance, fosfomycin resistance, beta-lactamase, cadmium resistance, multidrug resistance efflux pumps, and invasion and intracellular resistances. In the comparative eggNOG function study of six *S. aureus* genomes, "amino acid transport and metabolism" was observed as for most COGs, followed by those COGs associated with "translation, ribosomal structure, and biogenesis", "transcription", and "cell wall/membrane/envelope". The strain SA H32 genome occupied the highest number (2.16%) of COGs associated with "defense mechanisms (V)" (**Fig. 7b**). "Amino acid transport and metabolism (E)" was found highest (7.1%) in SA G8 genome. Functional annotation of the core-genome revelated that they are mostly associated with transcription and translation, and different metabolism categories, such a similar result was reported previously (Bosi et al., 2016; Sharma et al., 2018). The core-genome and accessory genome functional characterizations revealed that *S. aureus* isolates required amino acids rather than carbohydrates as the energy source and suggests that these isolates adapted to grow in a protein-rich medium than

carbohydrates **(Figs. 7a & 7b)**. It was suggested that the survival of *S. aureus* can be maintained by the catabolism of amino acids (Halsey et al., 2017).

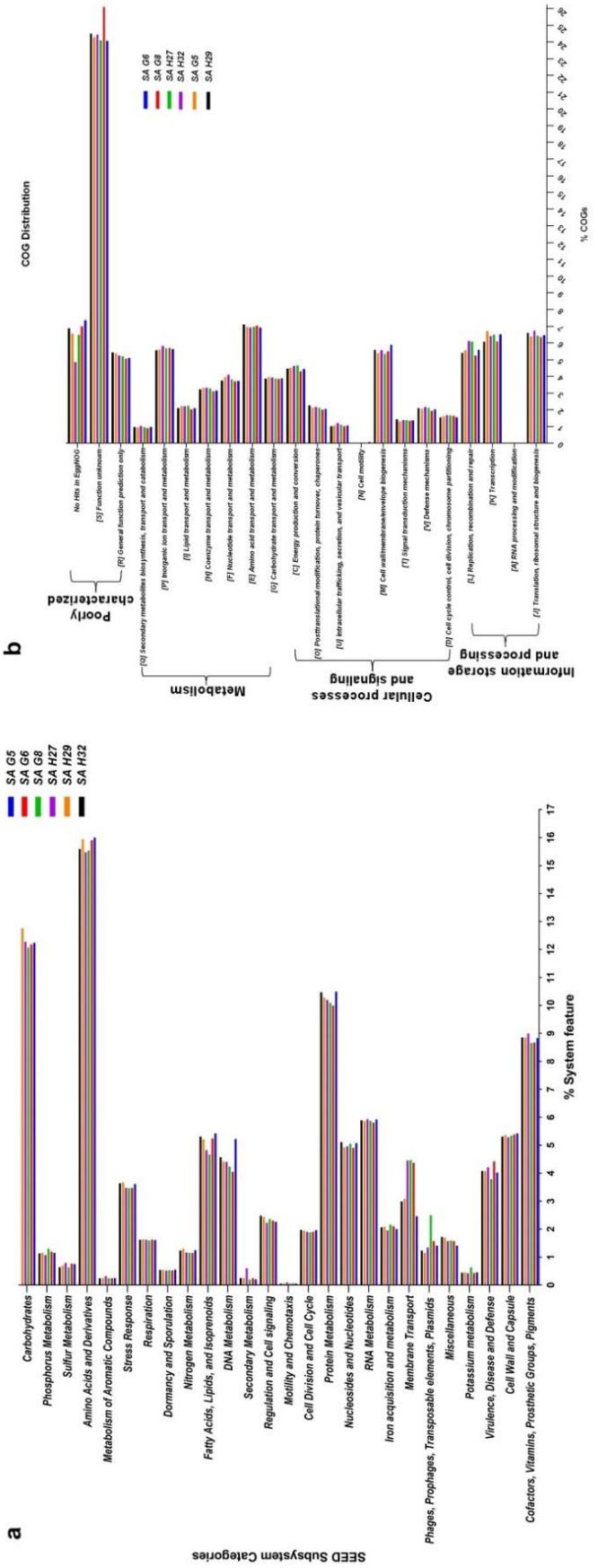


Fig. 7. Comparative functional categorization of all predicted ORFs in the genomes of the *S. aureus* isolates. (a) Percentage distribution of subsystem categories based on the SEED database. **(b)** Percentage distribution of COGs based on EggNOG.

5.10.2 Pan-genome, core-genome, and singletons analysis

The orthologous groups are categorized into three groups based on the pan-genome distribution such as core (present in all genomes of *S. aureus* strains), dispensable (present in at least two strains, but not all), and singleton genes (present no orthologs in any other genomes). The comparison of six study *S. aureus* genomes (SA G5, SA G6, SA G8, SA H27, SA H29, and SA H32) generated a pan-genome size of 3691 genes, of which 2202 (59.7%) genes were core-genome, 721 (19.5%) genes were dispensable, and 768 (20.8%) genes were singletons. The Rcp value for the genomes of *S. aureus* isolates was calculated and the ratio Rcp was 0.59 and it is indicated that the genomes of *S. aureus* isolates have a high degree of inter-species diversity and asymmetry (Ghatak et al., 2016). A total of 768 singleton genes were calculated across the genomes of six study *S. aureus* isolates, of which the SA G5 genome acquired the highest number of singleton genes (107) that constitute the genes encode for pathogenicity islands (SaPI), superinfection immunity protein, acetyltransferase (GNAT) family protein, type I restriction-modification system (DNA-methyltransferase subunit M, and specificity subunit S), Type III restriction-modification system methylation subunit, capsular polysaccharide synthesis enzyme (Cap8H, Cap8I, and Cap8J), alpha-aminoadipate--LysW ligase, O-antigen flippase, transcriptional regulators (MerR, TetR, and XRE), prophage-like elements, mobile elements, phage associated hypothetical proteins, hypothetical proteins, *etc*. The genome of SA G6 has 92 singletons that comprise the genes encoded for proteins *viz*. aminoglycoside 3-phosphotransferase, aminoglycoside 6-phosphotransferase, aminoglycoside N(6)-acetyltransferase, streptothricin acetyltransferase, antiseptic resistance protein, cadmium-transporting ATPase, mercuric ion reductase, anti-adhesin, Tn552 transposase, pathogenicity islands (SaPI), prophage-like elements, mobile elements, phage associated hypothetical proteins, hypothetical proteins, etc. The identified singleton genes of the SA G6 genome were present within the genomic island (GI). This GI region is located between 2804353–2873411 base pair sequence region of the genomic sequence. While the SA H27 genome has the least singleton genes (4) constituting the genes that encode for hypothetical proteins and phage proteins. The difference in the genomic constituents between the genome of SA H27 and SA H32 isolates revealed that SA H32 acquired the genes encoding for 23S rRNA (adenine (2058)-N (6))-dimethyltransferase, replication and maintenance protein, hypothetical proteins, phage-like elements, and mobile element protein. The genes

shared by six study genomes and their respective singletons are represented in **Fig. 8a**. When the nine reference *S. aureus* genomes (**Table S5**) were included in the pan-genome analysis, the core/pan-genome ratio drop down by 21.87% with inflation of pan-genome to 4443 genes and deflation of core-genome to 1679 genes. The core-genome and singleton genes formed by fifteen genomes of *S. aureus* strains are represented in the flower-plot **(Fig. 8b)**. When the nine reference genomes of *S. aureus* strains were included in the pan-genome analysis, SA G6 isolate occupied the highest number of singleton genes (46) while SA H27 isolate has the lowest singleton gene (1) among the present six study genomes (**Fig. 8b**).

In this pan-genome analysis of 15 *S. aureus* strains, the power-law coefficient, α value is 0.205 (between 0 and 1) which corresponds to the growing and open pan-genome model (Fitzgerald et al., 2001) **(Fig. S21a)**. The core-genome development plot showed the progression of pan and core-genomes as additional genomes are added for analysis and showed a sharp decline in core-genome size with the introduction of *S. aureus* subsp. *aureus* ST228 (HE579073) **(Fig. 8c)**. In the plot of core-genome development, the core-genome size approach (Ω) value revealed that the core-genome size of 15 *S. aureus* genomes would decline to 1541 **(Fig. S21b)**. The singleton development analysis suggested that the pan-genome size will continue to expand at the rate of 35 genes per novel, representative genomes **(Fig. S21c)**.

The shape of the pan-genome *vs.* core-genome curved showed fluctuation in their gene numbers when different order of the genomes was set, even so, pan-genome and core-genome development plots result remained unaffected by the genome order. The inflation of pan-genome and deflation of core-genome was seen after the introduction of reference genomes and its regression analysis revealed that the pan-genome is open, suggesting that the gene repertoire of this species is theoretically limitless. A similar finding was observed in the DNA microarray experiment of thirty-six *S. aureus* isolates (Nguyen & Kim, 2018; Schmidt & Hensel, 2004). The drastic decline of the core/pan-genome ratio after the introduction of HE579071.1 (*S. aureus* subsp. *aureus* ST228) and SA G6 suggested that these two strains have distinct genomic contents **(Fig. 8c)**. The genomic content variation between the genomes is due to the acquisition of certain genes that encode virulence and resistance factors, pathogenicity islands, prophage-like elements, plasmids, mobile element proteins, and hypothetical proteins in the GIs. These GIs are mobilized across organisms *via* HGT events (Hayek, 2013). This finding was supported by gaps that appeared in the genome

ring of SA G6 genome and suggested that this isolate showed a distant relationship to others **(Fig. 6)**.

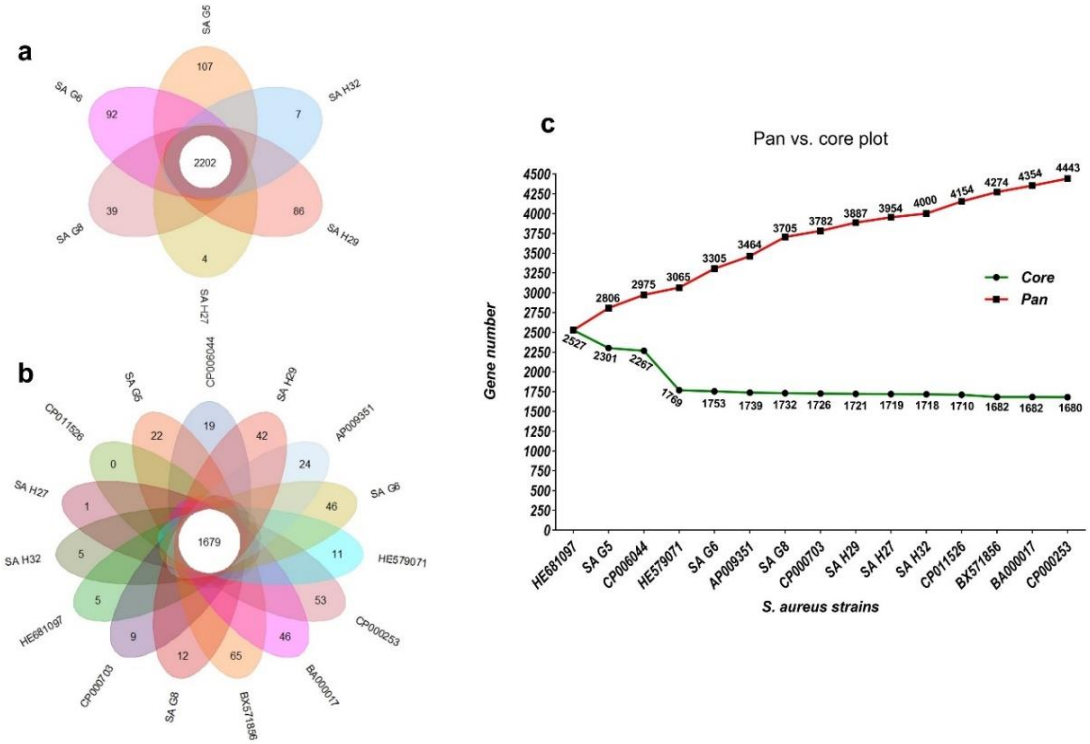


Fig. 8. Pan-genome analysis of *S. aureus* strains. The flowerplot diagram represents the core genomes and singletons of **(a)** six study *S. aureus* genomes, **(b)** six study isolates, and nine *S. aureus* reference strains. The core-gene was represented in the center of the flower and the petals represent the singletons of concerned genomes. **(c)** Core *vs.* pan-genome plot of the fifteen estimated *S. aureus* genomes.

The large repertoire of genes (62%) allocated in the accessory genome of 15 *S. aureus* strains gives advantages in adaptation and that can contribute to pathogenicity or niche specificity of strains (Ozer, 2018). The analysis of the pangenome is essential to understand the event of MGEs transfer and *S. aureus* evolution (Juhas et al., 2012). The interpretation from the dispersible and singleton genes content analysis of *S. aureus* genomes allows us to understand the genetic variation among the CC5, CC8, CC45, and CC22. Juhas et al. reported that most dispensable and singleton genes were acquired through HGT and operate an important role in drug resistance or virulence (Carvalho et al., 2019). A high portion of unique genes or singletons in *S. aureus* genomes were related to MGEs, which could drive the gaining of novel functional elements especially drug resistance and virulence. These singletons are the main

drivers of the phenotypic variation within *S. aureus* strains and the evolution of *S. aureus* (Aanensen et al., 2016).

5.10.3 Comparative phylogenetic tree analysis

The phylogenomic analysis of *S. aureus* isolates provides the tree into four major clusters **(Fig. 9a)**. Cluster A consists of 5 strains that belonged to CC22 and shows that Hungarian isolates, SA H27 and SA H32 have the highest proximity. Germany isolates, SA G6 and SA G8 isolate, and other strains belonging to CC5 were clustered in cluster B, showing that SA G6 isolate has closely relatedness to *S. aureus* subsp. *aureus ST228* (HE579071.1), and the SA G8 isolate has a higher relatedness to *S. aureus* subsp. *aureus* JH9 (CP000703.1) than SA G6 isolate. Hungarian isolate, SA H29 (CP032468), and other 4 strains that belonged to CC8 were grouped in cluster C. Also, a German isolate SA G5 (CP032160), and *S. aureus* CA-347 (CP006044.1) belonged to CC45 were grouped in cluster D. The phylogenetic relationship inferred from core-genome SNPs holds a similar agreement with the whole genome-based phylogenetic analysis, and these methods were able to distinguish between strains at a higher resolution in terms of the geographic origin of strains and phylogenetic trees are illustrated in **Fig. 9b**.

The phylogenomic analysis revealed that the strains with ST225 (Germany), ST228 (Germany, Switzerland), ST105 (USA), and ST5 (Japan) were grouped in the same CC5 cluster (Cluster B), and a different clade (Cluster A) was noticed among the UK origin ST22 (CC22) and diverged from Germany origin strains **(Fig. 9a)**, this finding was in good agreement with the previously published article (Denis et al., 2014). The CC5 (ST225) and CC22 (ST22) were found to be the most dominant clones circulating in Europe (Nübel et al., 2008, 2010; Vogel et al., 2012). The comparative genome analysis revealed that German and Hungarian isolates are genetically diverse and show variation among them due to the gain or loss of MGEs such as *SCCmec*, plasmid, phage elements, or the insertion of transposase. The event of MGEs transfer was observed in ST5, ST225, and ST228 **(Fig. 9a)**, and similar results were also reported previously (Pope et al., 2015).

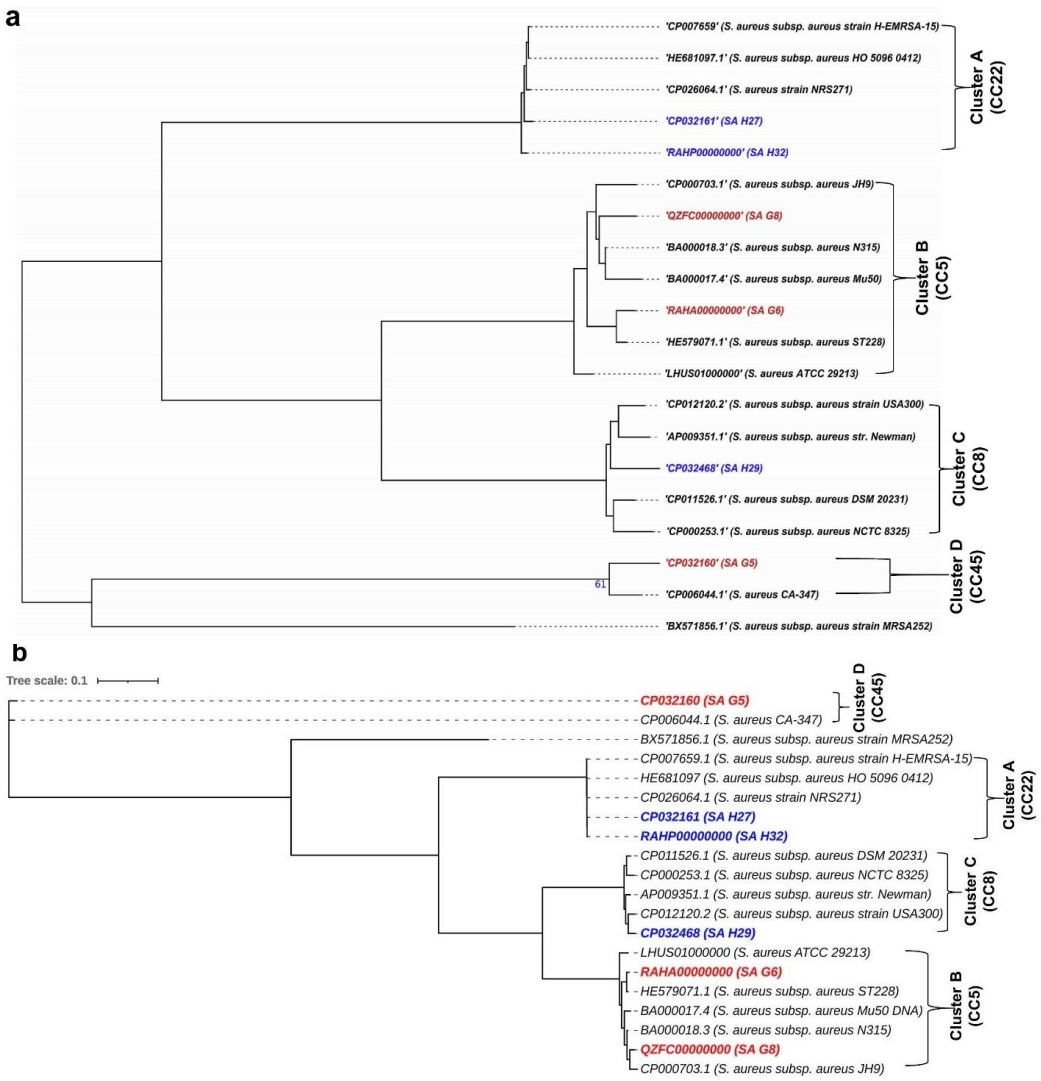


Fig. 9. Comparative phylogenetic analysis of *S. aureus* isolates strains with their closely related *S. aureus* strains. (a) Phylogenomic tree generated using genome sequences of *S. aureus*. The branch lengths are scaled in terms of GBDP distance formula d5. **(b)** Core-genome SNP tree was generated using an alignment of the high-quality SNPs and PhyML+SMS module was applied.

5.11 Retrieval of *S. aureus genome* sequences

To study prophages diversity, genome sequences of *S. aureus* strains associated with human (*Homo sapiens*) infections (n=34), bovine (*Bos taurus*) infections (n=22), and dog (*Canis lupus familiaris*) infections (n=4) were retrieved from NCBI database. The list of genomes and their respective features are shown in **Table S9**.

The transmission of *S. aureus* strains from animals to humans occurs commonly (Spoor et al., 2013). Animals-associated *S. aureus* strains may spread to the human population through various routes such as contact with contaminated meat products,

or infected farmers, butchers, and veterinary staff. Also, the contaminated effluent released from the animal farmhouses or veterinary hospitals could be another route for the transmission of *S. aureus* from animals to humans (Smith, 2015). The *S. aureus* with CC398 (ST398) and CC151 (ST151) are the most identified clone types of bovines in the European regions (Boss et al., 2016; Fluit, 2012). It was reported that these CCs are transmitted to humans and are considered to be an emerging zoonotic agent (Springer et al., 2009). In addition, the excessive or improper use of antibiotics in veterinary hospitals and animal husbandry promotes antibiotic-induced SOS response in *S. aureus strains* (Smith et al., 2002; Úbeda et al., 2005). This response triggers the phage induction and escalates the frequency of phage-mediated horizontal gene transfer (HGT) between the animals and humans-associated *S. aureus* strains (Goerke et al., 2006). For this reason, we selected prophages of CC398 and CC151 strains associated with animals to compare with prophages of CC398 or other CCs strains associated with human infections.

5.11.1 Identification and general genomic features of *S. aureus* prophages

PHASTER identified a total of 170 prophages of which 101 prophages (46 complete, 14 questionable, and 41 incomplete) were extracted from the genomes of *S. aureus* associated with human infections, 59 prophages (16 complete, 8 questionable, and 35 incomplete) were extracted from the genomes of *S. aureus* associated with *B. tourus* infections, and 10 prophages (3 complete, 6 questionable, and 1 incomplete) were extracted from genomes of *S. aureus* associated with *C. lupus familiaris* infections. The 65 intact/complete prophages were selected based on PHASTER scores (**Table S10)**. Among the 65 analyzed prophages, 57 prophages were extracted from methicillin-resistance *S. aureus* (MRSA) strains, while the 8 analyzed prophages were extracted from MSSA strains in which 4 prophages were from 4 MSSA strains (*S. aureus* HD1410, *S. aureus* I3, *S. aureus* SA13-192, and *S. aureus* SA14-639)-associated with human infections, and another 4 prophages were extracted from 4 MSSA strains (*S. aureus* 483, *S. aureus* 909, *S. aureus* C3489, and *S. aureus* C5086)-associated with bovine and dog infections. Among the MRSA strains, SA G6, and SA G8 strains were HA-MRSA, while the other MRSA strains were identified as CA-MRSA (Chua et al., 2014). The MLST analysis result revealed that *S. aureus* strains associated with the animal infections have ST398 (CC398) and ST151 (CC151), while the *S. aureus* strains associated with human infections have ST398 (CC398), ST8

(CC8), ST5 (CC5), *etc.* The genome sizes of intact/complete prophages were in the range of approximately 24.8 to 87.8 kb, and the GC content varied between 32.16 and 35.38%. These complete prophages belonged to the *Siphoviridae* family and had temperate lifestyles. The highest number (132) of coding sequence (CDS) was found in phiG4-3 (*S. aureus* SA G6) and the lowest number (29) of CDS was found in phiH7-2 (*S. aureus* strain H7) **(Table S10)**.

5.11.2 Sequence clustering and phylogenetic relationship of the prophages

Sequence clustering was performed by aligning the whole-genome sequences of 65 intact/ complete prophage genomes carried by human and animal-associated *S. aureus*. The generated phylogenetic tree grouped the prophages into 7 different clades **(Fig. 10, & Table S11)**.

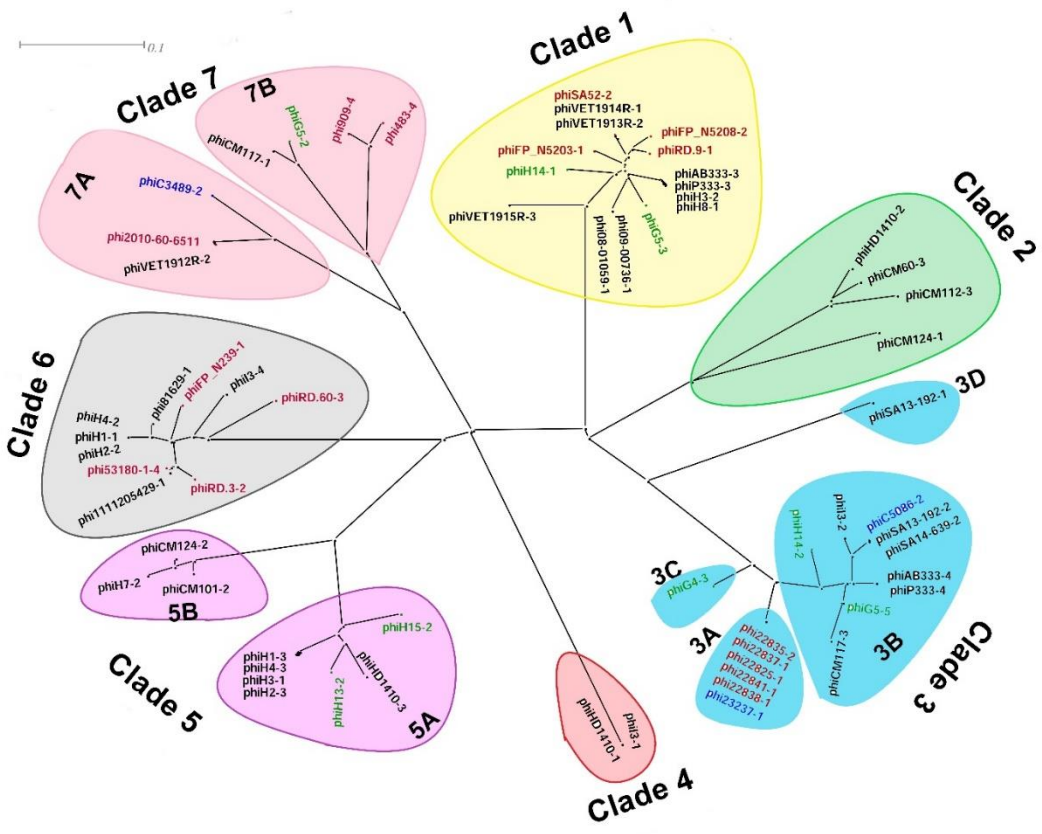


Fig. 10. Cluster generated from the prophage genome sequences and phylogenetic relationship. The prophage genome sequences were aligned using MAFFT tool with default parameters, and the aligned sequences were used to create phylogenetic tree by the BioNJ algorithm in SplitsTree4 software. The clades were shaded with different colors and the prophage names labeled in black, red, and blue colors represent the prophages extracted from the *S. aureus* associated with humans,

bovines, and dogs, respectively. And the prophage names labeled in green color represent the prophages extracted from newly sequenced genomes.

The prophages carried by human and animal-associated *S. aureus* strains linked with different infected sites and different geographical locations were dispersed randomly in all clades. Although it was expected that the prophages of animal-associated *S. aureus* and human-associated *S. aureus* strains would form individual clusters, this was not the case. There was no correlation between the clade, the host, the site of infection, or geographical locations. This analysis showed that the prophages of the same host were dispersed in different clades rather than appearing in a single clade **(Fig. 10)**. This finding suggested each *S. aureus* strain carried two or more different prophages with unique features. It was reported that a bacterial cell owning one or more prophages is considered as a lysogen that provides immunity toward the infection by the same group of phages (Clokie et al., 2011) The phylogenetic analyses result showed no difference in clustering patterns of prophages carried by HA-MRSA and CA-MRSA strains. The prophages (phiG5-2, phiG5-3, phiG5-5) carried by the HA-MRSA strain were clustered with other prophages of CA-MRSA strains in clade 1, subclade 3B, and subclade 7B. Besides, the prophages carried by MSSA strains showed high proximity among them and found clustered in subclade 3B (phiI3-2, phiC5086-2, phiSA13-192-2, and phiSA14-639-2), clade 4 (phiI3-1, and phiHD-1), clade 7 (phi483-3, phi909-4, phiC3489-2, and phiC5086-2) **(Fig. 10)**. This similar clustering pattern and prophages sequences similarities are favored by the identical *SCCmec* types, STs, or CCs of their host strains. This study suggested that the frequency of phage-mediated HGT is higher between the *S. aureus* strains with the identical *SCCmec* type, STs, or CCs.

5.11.3 Comparative genome analyses of *S. aureus* prophages

Prophages are a part of the accessory genome in a bacterial genome; however, identified prophages themselves have a pan-genome of 107,158 bp size. Notably, identified prophages did not have a core-genome that is conserved among all prophages across their phylogeny **(Fig. S22)**; such a similar finding was reported previously (Lucchini et al., 1999). The presence of functional modules with low sequence similarity may be due to the recombination of two or more prophages within host genomes or the horizontal exchange of functional modules between related phages (Pope et al., 2015). Furthermore, the presence of variable MGEs, bacterial

genes, and unspecified genes in the genome of prophages which were thought to be acquired from the different *S. aureus* strains suggested that such prophages had undergone several HGT events which result in prophage genomes with high variation (Cumby et al., 2012), and rapid emergence of new phages (Kaneko et al., 1998; Narita et al., 2001). The low sequence similarity in the identified prophage genomes made it difficult to generate their core-genome. To overcome this limitation, we performed clade/subclade-wise prophage genomes analyses based on gene-by-gene alignment at a finer synteny level.

Clade 1 comprises 15 prophages with a total sequence size of 122,268 bp encoding 168 CDS features. The sequences share 19.84% similarities, and the similar region encodes 28 common CDS features **(Table S11)**. In this clade, phiSA52-2 showed distinctive CDS features and possess the highest accessory genes (67 CDS) while phiH8-1 has low accessory genes (30 CDS) **(Fig. 11a-b & clade 1 of Fig. S23)**. The prophages carried by *S. aureus* associated with human or bovine infections have relatively high genome sizes in comparison with prophages of *S. aureus* associated with dogs.

In **fig. 11a** synteny, phiAB333-3, and phiP333-3 showed the highest sequence identities, and their host *S. aureus* strains owned *SCCme*c IVc type and ST8 and belonged to the same geographical origin (France), but these prophages were carried by *S. aureus* associated with human skin and nares infections. respectively. And prophage names labeled in green color indicated the prophages extracted from the study genome and grey shaded regions are homology regions. Similarly, in **fig. 11b** synteny, phiVET1913R-2, and phiVET1914R-1 showed the closest relationship in this clade, and these two prophages were from the same country of origin (Netherlands) and their host *S. aureus* strains carried *SCCmec* Vc type and ST398, however, they were found in the *S. aureus* associated with the throat and nasal infections. Clade 2 has 4 prophage sequences which showed a total sequence size of 112,518 bp encoding 149 CDS features. These sequences shared 6.11% bp similarity that encodes 7 common CDS features **(Table S11)**.

In this clade, phiCM124-1 showed distinctive features compared with the other prophage sequences and possessed the highest accessory genes (80 CDS), however, phiCM112-3 has the lowest accessory genes (28 CDS).

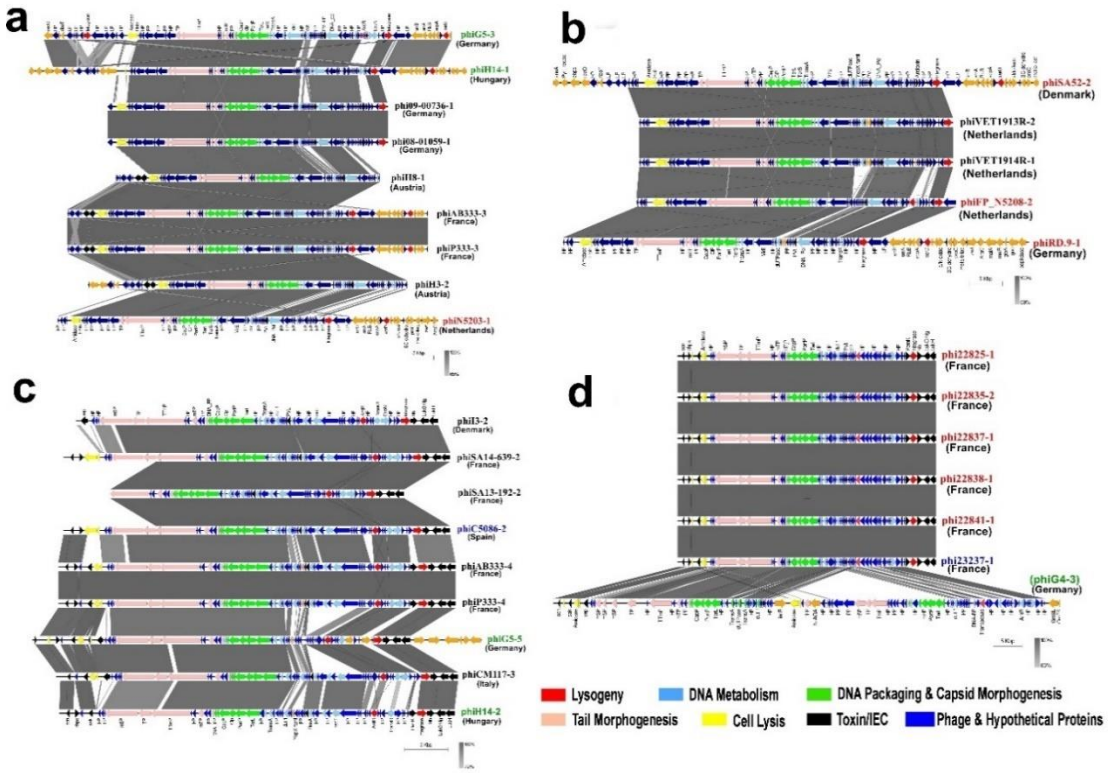


Fig. 11. The mosaic architecture of *S. aureus* prophage genomes.

(a) Comparison of prophages of clade 1, representing phiG5-3, and phiH14-1 sequenced of the current study and 7 prophages associated with *S. aureus* reported from Germany, Austria, France, and the Netherlands; **(b)** Comparison of prophages of clade 1, representing prophages of *S. aureus* associated with humans, and bovines; **(c)** Comparison of prophages of clade 3, representing prophages phiG5-5 & phiH14-2 (subclade 3B) sequences and 6 prophages associated with *S. aureus* reported from Denmark, France, Spain, Germany, Italy, and Hungary; **(d)** Comparison of prophages in clade 3, representing prophages phiG4-3 (subclade 3C) sequence and 6 prophages (subclade 3A) associated with *S. aureus* spreading in France, Spain, and Germany. Phages and country of origins are indicated on the right. Prophage names labeled in black, red, and blue color indicated prophages associated with humans, bovines, and dogs, respectively, and the prophage names labeled with the green color indicated the prophages extracted from the six study *S. aureus* genomes.

In clade 3, the subclade 3A consists of 6 prophage sequences showing 100% bp sequence similarity to each other which encodes 69 CDS features, however, phi23237-1 was from a different host spectrum **(Fig. 10, Table S11 & Fig. 11d)**. The **fig. 11d** synteny revealed the highest degree of sequence similarity between these prophages. The host of these prophages owned the identical *SCCmec* IVa /ST398. Remarkably,

this showed the evidence of phage-mediated horizontal gene transfer between the *S. aureus* associated with bovine infections and the *S. aureus* associated with dog infection **(subclade 3A, & fig. 11d)**. The 9 prophage sequences of the subclade 3B have a total sequence size of 74,025 bp that encode 111 CDS features. It showed 27.23% bp similarity that encodes 29 common CDS features **(Fig. 10, Table S11 & Fig. 11c)**. The highest accessory genes of 40 CDS were observed in the phiG5-5 while, phiSA13-192-2 occupied the lowest accessory genes (24 CDS) **(clade 3B of Fig. S23)**. F**ig 11c** synteny revealed that phiAB333-4 and phiP333-4 showed 100% sequence similarities in which their hosts carried the same *SCCmec* IVc type and ST8. The phiG4-3 and phiSA13-192-1 were distinct from other subclades 3A, and 3B resulting in discrete subclades 3C, and 3D, respectively **(Fig. 10 & Table S11)**. In clade 4, phiI3-1, and phiHD1410-1 have a total sequence size of 33,850 bp and they share 100% sequence similarity although they have different host spectra. However, their host strains were identified as MSSA and owned the same CC30 **(Fig. 10 & Table S11)**. In clade 5, subclade 5A encompasses 7 prophage sequences with 119,097 bp total sequence size which encodes 159 CDS features **(Fig. 10)**. The prophage sequences in this subclade have 17.64% sequence similarities and share 23 CDS features **(Table S11)**. In this subclade, the highest accessory genes were found in the phiH3-1 (75 CDS), and the lowest in phiHD1410-3 (59 CDS) **(clade 5A of Fig. S23)**. In this subclade, phiH1-3 and phiH2-3 showed the highest sequence identity, and phiHD1410-3 has closed sequence similarity with the reference phiNM-3 **(Fig. 11a)**. Subclade 5B consists of 3 prophage sequences that shared 30.36% sequence similarities out of a total sequence size of 77,303bp and have 27 common CDS features **(Fig. 10 & Table S11)**. The phiH7-2 sequence has the lowest number of accessory genes. However, the phiCM124-2 and phi101-2 showed a high similarity of 54.02% base-pairs and shared 50 CDS features.

Clade 6 has 10 prophage sequences with a total sequence size of 133,139 bp that encodes 170 CDS features **(Fig. 10 & Table S11)**. It showed 20.81% sequence similarity that encodes 29 CDS features **(Fig. 12 & clade 6 of Fig. S23)**. The phiRD.60-3 has the highest number of accessory genes (57 CDS) while the phi81629-1 has the lowest number of accessory genes (5 CDS). The phiH4-2, phiH1-1, and phiH2-2 shared 99.77% sequence similarities and have 93 common CDS features **(Fig. 12b)**.

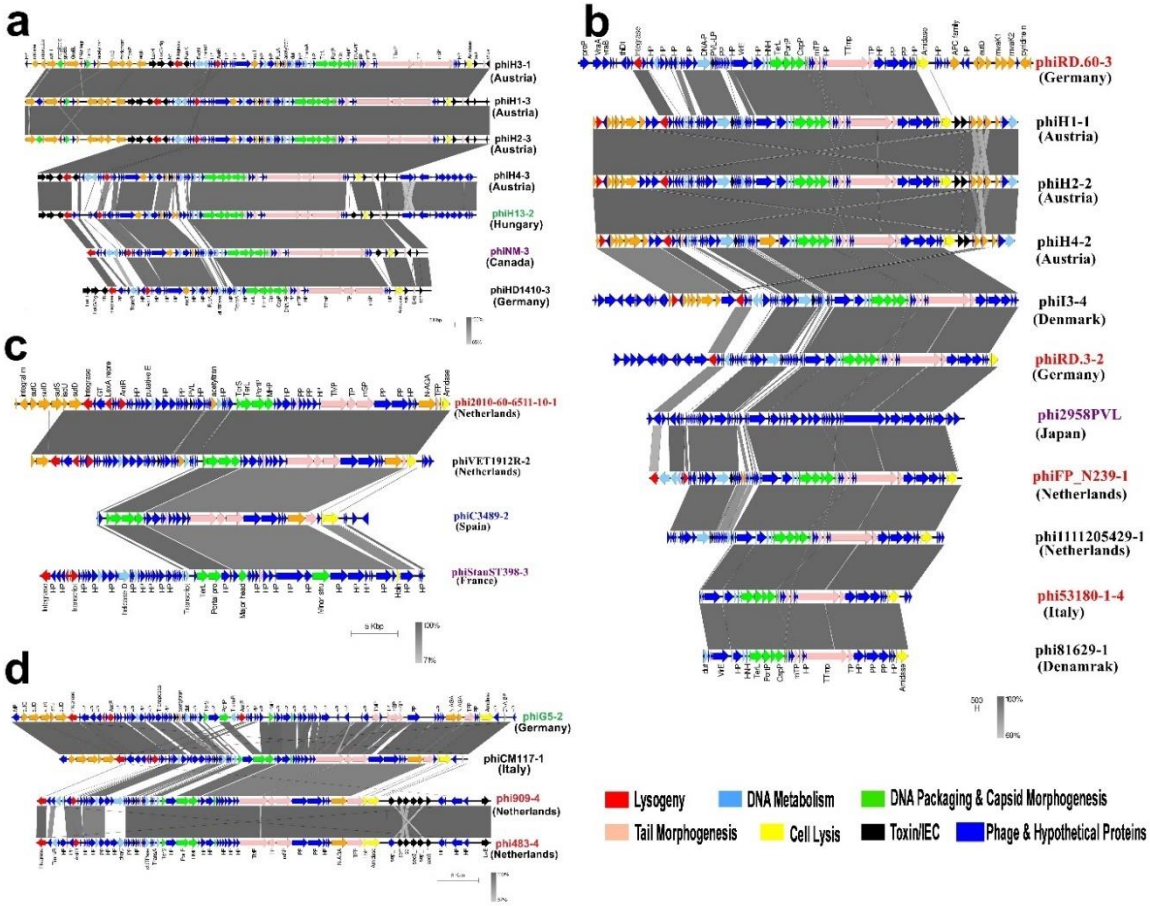


Fig. 12. The mosaic architecture of *S. aureus* prophage genomes.

(a) Comparison of prophages of clade 5, representing the prophages of subclade 5A and phiH13-2 sequenced reported from Austria, Hungary, Canada (reference prophage), and Germany; **(b)** Comparison of prophages of clade 6, representing prophages of humans and bovines -associated *S. aureus* spreading in Germany, Austria, Denmark, Netherlands, Japan (reference prophage), and Italy; **(c)** Comparison of prophages of clade 7, representing subclade 7A prophages humans, bovines, and dogs-associated *S. aureus* reported from Netherland, Spain, and France; **(d)** Comparison of prophages of clade 7, representing phiG5-2 prophage sequence and 3 prophages of subclade 7B associated with *S. aureus* reported from Italy, and Netherland. Phages and country of origins are indicated on the right and the prophage names labeled in black, red, blue, and purple color indicate prophages associated with humans, bovines, dogs, and references, respectively. And prophage names labeled in green color indicated the prophages extracted from study genomes and grey shaded regions are homologous regions.

In **fig. 12b** synteny, the host strains of prophages (phi53180-1-4, and phi81629-1) were associated with bovine, and human infections, and occupied distant geographical locations (Italy, and Denmark), however, the host strains of these prophages carried the same *SCCmec* IVa type, and ST398 (CC398), as a result, their prophages revealed high sequence similarities. Moreover, phiFP_N239-1 and phi2958PVL showed high sequence identity **(Fig. 12b)**.

In clade 7, the subclade 7A contained 3 prophage sequences with 56,344 bp total sequences that share 38.03% sequence similarities which encode 25 CDS features out of 86 total CDS features **(Fig. 10, & Table S11)**. The phiC3489-2 displayed the lowest number of CDS features (10), while the phi2010-60-65511-10-1 showed the highest number of CDS features (48) **(Fig. 12c, & clade 7A of Fig. S23)**. The reference phiStauST398-3 showed a high sequence identity with phiVET1912R-2 **(Fig. 12c)**.

Subclade 7B consists of 4 prophage sequences having 98,346 bp that codes for 115 CDS features. This subclade shared 26.35% similarities that encode 45 CDS features **(Fig. 10, Table S11, Fig. 12d, & clade 7B of Fig. S23)**. Glycosyltransferase family protein was detected in phi2010-60-65511-10-1 and phVET1912R-2 prophage sequences. This protein performs the conversion of host serotype during lysogeny of temperate phage (Gill et al., 2005; Lee & Iandolo, 1986). In clade 7B, the *S. aureus* strains isolated in the Netherlands that were associated with bovine milk were identified as MSSA strains and both have the same ST151, as a result, their prophages showed high sequence similarities.

5.11.4 Putative virulence factors associated with *S. aureus* prophages

The prophages of *S. aureus* associated with human and animal infections were found to harbor the virulence factors encoding genes that function in immune evasions, tissue evasions, toxins, adherence, and iron uptake. The comparative analyses of virulence factors encoding genes associated with prophages are summarized in a heatmap **(Fig. S24)**. The highest prevalence of toxin encoding genes was observed in clade 7. The lowest prevalence of virulence encoding genes was observed in clades 2 and 4.

The prophages belonging to clade 1, 6, and prophage phiH15-2 of clade 5 showed the presence of *clpP* and *virE* genes. Some of the prophages of clade 1 also showed the presence of *lukF-PV* and *lukS-PV*, *yopX*, *sak*, *scn*, and *chp*. Besides, *hlb*, *ebp*, and

xerD genes were also detected in the prophages of Clade 1. The prophages phiCM124-1 of clade 2 showed the presence of *isd* gene clusters (*isdA-isdH)*. The genes carried by the prophages of clade 3 mostly include *hlb, sak, scn,* and *chp*. In addition, *yopX* gene and genes that encode enterotoxins, *sep,* and *sec* were also detected in some prophages of clade 3. The *xerD* gene was also carried by the prophages phiH4-2, phiH2-2, phiH1-1, and phiI3-4 of clade 6, and clade 4. The prophages of clade 5 carried genes such as *sak, scn, chp,* and *hlb*. However, some prophages also showed the presence of a *sea* gene. *yopX* and *eap*/*map* genes. The prophages of clade 6 showed the presence of *lukF-PV* and *lukS-PV, clpP, xerD, virE,* and *yopX* genes. However, prophage phiRD.3-2 carried *geh* gene encoding lipase protein which might be a virulence factor and associated with the lysogenic conversion of *S. aureus* (Kala et al., 2014). Most of the prophages especially phi909-4 and phi483-4 of clade 7 carried the toxin encoding genes *viz*., leukocidin-related gene (*lukE*), and enterotoxin genes (*seg, sei, sem, sen,* and *seo*). These prophages have genes that encode the HNH endonuclease, a key component of phage DNA packaging machines (Feng et al., 2008; Gordon & Lowy, 2008; Tenover & Gaynes, 2000), and VRR-NUC domain protein (virus-type replication-repair nuclease). The variation in virulence factor encoding genes in prophages within each clade and among clades showed the genomic diversity of prophages, evolution, and the emergence of highly pathogenic *S. aureus* strains.

The highest prevalence of immune evasion encoding genes was observed in the prophages of clades 3 and 5, however, such genes were not present in the genomes of prophages of clades 6 and 7 **(Fig. S24)**. The prophages carried by animal-associated *S. aureus* have 50.34% more VFGs than the prophages harbored by human-associated *S. aureus*. This observation suggests that the prophages carried by animal-associated *S. aureus* may have more pathogenic potential. Leucotoxin (*lukE*) was found in 2 prophages harbored by animal-associated *S. aureus.* The PVL toxin was found in the prophages (phiAB333-3, phiP333-3, phiH8-1, phiH3-2, and phiG5-3) of clade 1, and (phiH4-2, phiH2-2, and phiH1-1) of clade 6. This toxin has been implicated in the pathogenesis of severe necrotic infections of higher vertebrates (van Wamel et al., 2006). These PVL-positive prophages were carried by human-associated *S. aureus* strains. These human-associated *S. aureus* strains were belonged to CA-MRSA, except the phiG5-3 in which its host strain belonged to HA-MRSA. This result indicated that the PVL-positive prophages have a lineage-specific relationship (Boakes et al., 2011). Nevertheless, such PVL-positive prophages showed mosaic

genomic architectures, and their host strains were from different geographical locations. These changes in genetic sequence and geographical variations play a significant role in the evolution of MRSA clones and offer an important insight into the microepidemiology of PVL-MRSA (Coombs et al., 2020; Waldron & Lindsay, 2006). The PVL encoding genes were located between the phage lysin and the *attR* site and the phages showed an elongated head instead of icosahedral (Narita et al., 2001). It has been reported that helper phage φSLT mediates the transfer of these genes from a PVL-positive to a PVL-negative *S. aureus* strain during positive lysogenic conversion (Deghorain & Van Melderen, 2012). The acquisition of PVL-positive phage by epidemic HA-MRSA strains could surge morbidity and mortality in the case of nosocomial MRSA infection (Boakes et al., 2011).

In the present study, IEC (*sak, chp,* and *scn*) was identified in prophages harbored by both human and animal-associated *S. aureus* which are known for positive lysogenic conversion. This IEC can be easily transferred from one *S. aureus* strain to another by a diverse group of *hlb*-converting bacteriophages and contributes to the pathogenomic diversity and human niche-specific adaptation of *S. aureus* strains (Verkaik et al., 2011). The IEC is highly human-specific, and it is assumed that the IEC-containing phages are less prevalent in animal isolates (McCarthy, Witney, et al., 2012; Price et al., 2012; Resch et al., 2013) and are lost when *S. aureus* shifts its host from human to animal (Fortier & Sekulovic, 2013). The disruption of chromosomal factors through phage integration was known as a lysogenic negative conversion and resulted in the inactivation of *geh,* and *hlb* genes (Lee & Iandolo, 1986). The prophage, phiRd.3-2 of *S. aureus* associated with bovine infection possesses *geh* and it was reported that the integration of the *geh* gene causes negative lysogenic conversion in Staphylococcal phage L54a (Nowrouzian et al., 2015). Our study observed the high prevalence of enterotoxin-gene cluster (*egc*) such as *seg, sei, sem, sen, seo,* and *seu* in the prophages of animal-associated *S. aureus.* This *egc* was identified mainly in phi483-3, and phi909-4 of clade 7 harboring the novel SaPIs **(Fig. S24)**. It was reported that *egc* acts as a colonization factor thereby magnifying the commensal fitness and showing aggravating effects in bacteremia (Xia & Wolz, 2014). Besides, enterotoxin-encoding genes (*sea* and *sep*) were found in the prophages carried by human-associated *S. aureus* **(clade 5 of Fig. S24)** which are responsible for causing infections such as food poisoning, toxic shock syndrome, necrotizing fasciitis bullous impetigo, and chronic bovine mastitis (Reniere & Skaar, 2008). The prophage

phiCM124-1 showed the presence of *isd* gene clusters which are required by the *S. aureus* for iron acquisition and resulting in lysis of erythrocytes in humans and animals (Hussain et al., 2008). Also, extracellular adherence protein encoded by *eap* gene observed in phiH14-1, phiH3-2, is reported to facilitate adherence of *S aureus* to host extracellular matrix components and prevents inflammation, wound healing, and angiogenesis (Baumler & Fang, 2013). Besides, the presence and distribution of different virulence factors encoding genes among the prophages of various clades suggested that either the presence of these factors is dependent on the host environmental condition, or these factors allow the host bacteria to adapt to different environmental niches (Tang et al., 2013).

The prophages carried by human-associated *S. aureus* strains of different serotypes obtained from different geographical locations scattered themselves in all the clades, suggesting that these phages have a wide distribution across the European regions. A similar finding was also reported earlier in prophages of *Streptococcus suis* (Tang et al., 2013). The prophages carried by animal-associated *S. aureus* also showed their presence in different clades *viz*., clade 3A, clade 6, and clade 7B **(Fig. 10)** and displayed slightly different clustering patterns compared to the clustering pattern of human-associated *S. aureus* prophages. Overall, this study demonstrates that the presence of prophages in the genome of *S. aureus* associated with both humans and animals causes genetic variations in the bacterium, confers antibiotic resistance, and helps the bacterium to adapt to hostile conditions and that in turn increases its pathogenicity. The lateral transfer of genes encoding ARGs by phage-mediated transduction could be an important contributing factor in the global spread of antibiotic resistance. In this study, all the CC398 strains were identified in MRSA strains and showed high prevalence in animal-associated *S. aureus* strains. The prophages carried by CC398 clone of animals and humans associated with *S. aureus* strains showed dispersed in different clades **(Fig. 10)**. The presence of similar genetic elements in the prophages isolated from *S. aureus* associated with animals and humans suggested that prophages may have played a major role in the epidemiological changes. The appearance of the mosaic nature of prophage genomes suggested the occurrence of genetic exchange among the *S. aureus* strains *via* phages. Also, the presence of VFGs in the genomes of prophages supports *S. aureus* to adapt to different environmental niches, promote the pathogenesis and facilitate their evolution. The IEC was identified in both prophages harbored by human and animal-associated *S.*

aureus which are a human niche-specific adaptation of *S. aureus* strains. The IEC is highly human-specific, however, our findings revealed that the presence of IEC could not differentiate between phages of human and animal-associated *S. aureus.* The presence of various virulence factors in the genomes of prophages of animal-associated *S. aureus* suggested that these prophages could have more pathogenic than the prophages of human-associated *S. aureus.* Comparative studies of prophages carried by human and animal-associated *S. aureus strains* have very crucial importance for the investigation of *S. aureus* transmission from human to animal and *vice-versa*, as well as to gain a better understanding of their evolutionary relationships, and diversity.

5.12 Identification and characterization of potential drug and vaccine target candidates by reverse vaccinology

This study was to identify the new drug and vaccine targets against MRSA strains for proposing alternative prospective therapeutics. Earlier, the development of new antimicrobial agents and vaccine therapies was limited due to a computational technology bottleneck. However, in this post-genomic era, advances in the fields of genomics, and proteomics, coupled with the development of bioinformatics tools have allowed for in-silico identification of new drug and vaccine targets from genomic, and protein sequence resources using the subtractive genomic approach or reverse vaccinology approach (Solanki et al., 2019). The reverse vaccinology approach is a powerful method of identification of unique yet uncharacterized sequences as possible therapeutic targets (Amineni et al., 2010; Khalida et al., 2012; Reddy et al., 2010). Reverse vaccinology reduces the time needed to develop new vaccines and allows vaccines to be designed even for non-cultivable pathogens (Gupta et al., 2020). This approach also allows not only the detection of all the antigens as observed through the conventional methods but also the discovery of novel antigens that function in a different paradigm (Rappuoli, 2000).

5.12.1 Prediction of core-proteome

A set of 1719 coding DNA sequences (CDSs) shared by 16 MRSA genomes (Table S5) were identified and these sequences were extracted in form of protein sequences and considered as core-proteome. Out of this core-proteome, 1678 non-redundant protein sequences were retrieved by using CD-HIT with tolerance at 70% threshold.

5.12.2 Identification of essential proteins and non-homologous proteins

The non-redundant protein sequences of the core-proteome have large numbers of genes that are not essential for the survival of an organism. Essential proteins show potential targets for drug designing because the mainstream antibacterial compounds are synthesized to dock essential proteins. This step primarily mines essential proteins of MRSA strains that are necessary for their survival within the host. Such proteins are housekeeping in nature and important for basic cellular functions (Zhang & Ren, 2015). If essential proteins are functionally classified as virulent, they are unique and of vital significance, in unveiling novel therapeutic targets as these proteins help bacteria to modulate or reduce host defense mechanisms and may promote pathogenesis (Barh et al., 2011). The essential core-proteome that has homology among all the *S. aureus* strains has the potential to serve as a drug/vaccine target to combat multi-drug resistant *S. aureus* strains (Uddin & Sufian, 2016). Essential core proteins of 278 were identified by GEPTOP 2.0 server. Further comparison of the core-proteins of essential protein sequences to the human host proteome resulted in a set of 98 targets as essential non-host homologous and a set of 184 targets as essential host homologous proteins. Proteins that are non-homologous and essential for pathogens hold great promise to develop species-specific potential drug targets (Goncearenco et al., 2017), and this subtractive proteomic or genomic approach is very essential to avoid drug cross binding with the human host proteins and the possibility of the drug adverse effects (Uddin et al., 2019).

5.12.3 Characterization and prediction of subcellular location of proteins

The conserved non-host homologous proteins of MRSA strains can localize at different regions including cytoplasmic, membrane, putative surface-exposed, and secretory. Therefore, the localization of proteins is an important aspect of designing any therapeutic agents such as drug targets or vaccine candidates (Goyal & Singh, 2018). Out of the 98 subcellular localized and essential non-homologous proteins, 78 proteins were localized in cytoplasmic (CYT), 2 proteins were secretory (SEC), 6 proteins were potentially surface exposed (PSE), and 12 proteins were membrane (MEM) bound. The proteins localized in cytoplasmic regions play a pivotal role in maintaining cell viability and therefore, these proteins are considered as drug targets (Duffield et al., 2010). The exposed proteins including membrane, putative surface exposed, and secreted proteins are better applied as vaccine targets for reverse

vaccinology (Rappuoli, 2001). The membrane proteins are more priority for vaccine candidates due to their closer contact with the host and cause the immune responses, however, these proteins with more helices are difficult to purify from the bacteria, thus membrane proteins containing fewer helices (≤2) are more preferred (Mondal et al., 2015).

5.12.4 High-throughput structural modelling and druggability analysis

The CYT protein sequences were selected for drug target analysis using MHOLline 2.0, an online web tool, to predict the modelome. In this analysis, proteins that belonged to very high, high, and good structural qualities of the G2 model group was considered for further analysis. This G2 model predicted 19 proteins (9 very high quality, 8 high, and 2 good) as potential candidates for drug targets. The other factors involved in drug target prioritization are low molecular weight, and high druggability (Abadio et al., 2011). The ability of a protein to hold a pocket for the binding of small molecules is one of the key steps in the identification of a drug target (Hussein et al., 2017). Therefore, pocket druggability analyses are crucial in therapeutic drug discovery. In this study, we chose the 8 best potential drug target candidates based on the drug score de-fined by the DoGSiteScorer tool, and a molecular weight criterion of less than 90 kDa **(Table S12)**.

5.12.5 Virtual screening and molecular docking analysis

The ligands, drug-like molecules were screened for favorable interactions with each target protein. The top compounds were further used for flexible docking analysis with the residues of the most druggable cavity defined by DoGSiteScorer and MVD software. As the result, the predicted protein-ligand interactions with the active site residues of each target are represented in **Table 3**, with ZINC ID, MolDock score for the selected ligand as well as hydrogen (H)-bonds involved in the interaction. In molecular docking, lower energy scores represent better protein-ligand bindings compared to higher energy values (Thomsen & Christensen, 2006). Hydrophobic interactions are the major contributors to the stability of proteins. H-bonding also maintains protein stability, but to a lower extent than hydrophobic interactions, even in the smallest globular proteins. Accordingly, hydrophobic binding of a ligand to essential amino acid residues of protein is the main determinate of folding configuration equilibria in many native proteins (Pace et al., 2011).

Table 3. Docking studies of drug-like molecules (ZINC Compounds) with eight drug target proteins. The table shows the MolDock score, number of hydrogen bonds, and active site residues of target proteins with the respective ZINC compounds.

Target Proteins	ZINC ID	MolDock Score	H-bonds/ Residues
Biotin protein ligase	ZINC4235426	-176.846	7/ Tyr182, Arg227, Arg125, and Arg122
Thymidylate kinase	ZINC4259578	-139.656	3/ Arg75, Arg97, and Arg110,
	ZINC4235426	-139.150	6/Arg75, Arg97, Glu106, Tyr105, and Glu42
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase	ZINC4235426	-125.654	5/ Tyr45, Thr46, Val47, and Glu155
Phosphate acetyltransferase	ZINC4270981	-134.847	2/ Gln325, and Leu299
HPr kinase/phosphorylase	ZINC4235426	-147.451	4/Lys259, Thr150, and Asn227
	ZINC4235924	-137.549	7/Gly151, Thr150, Asn227, Lys258, and Asn229
Phosphate acetyltransferase	ZINC4270981	-134.847	2/ Gln325, and Leu299
UTP-glucose-1-phosphate uridylyltransferase	ZINC428871	-122.664	1/ Leu110
	ZINC31154666	-120.197	4/ Gly171, Leu230, Leu110, and Asn226
Fatty acid/phospholipid synthesis	ZINC4237105	-130.756	1/ Lys262
Pantothenate synthetase	ZINC4235426	-173.843	4/ Met31, Gly148, His35, and Thr30

5.12.6 Biotin protein ligase

Biotin protein ligase (BPL) plays two important roles such as biotinylation, and transcription repressor activities that involve in biotin homeostasis (Soares et al., 2013). This enzyme plays the key master regulator of all biotin-mediated metabolic processes in *S. aureus* and is an emerging new drug target (Soares et al., 2013; Feng et al., 2016). It is also reported that this enzyme is essential for fatty acid biosynthesis and the tricarboxylic acid cycle pathways in *S. aureus* (Soares et al., 2013). The crystallographic structure of the BPL template (PDB ID: 6NDL from *S. aureus*) has two crystallographic native ligands, where BQX ligand participating in H-bond interaction on active site residues of its template were Asp180, Arg125, Ser128, Arg120, Arg122, Lys187, Gln116, Ser93, Asn212, and Thr94. In this study, the cavity 1 defined by the MVD tool having Volume (V):160.768 Å; Surface (S):528.64 Å; Radius (R):20 Å of BPL template was docked with ZINC4235426 compound resulting in the formation of 7 H-bonds with the active site residues, Try182, Arg227, Arg125, and Arg122 and shown a MolDock score of -176.846 (**Table 3, Fig. 13**). Also, the ZINC4237101 compound formed no H-bonds with the active site residues of the BPL template and revealed the MolDock score of -167.239.

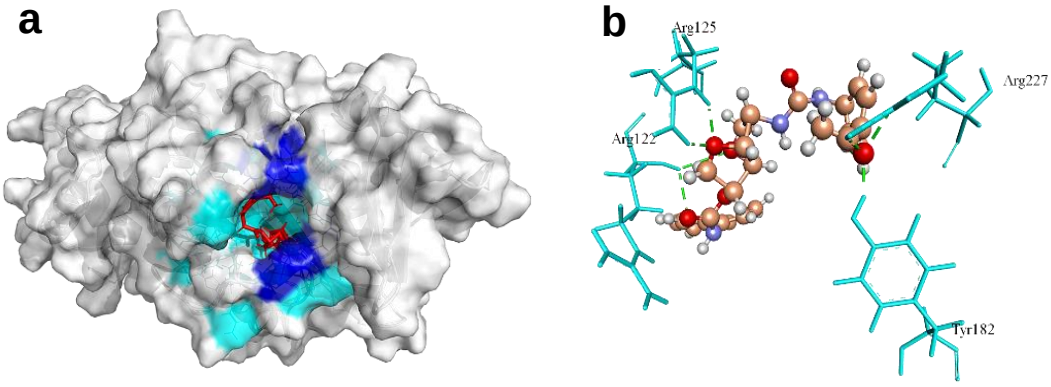


Fig. 13. The molecular docking analysis of BPL (WP_000049913.1) with compound ZINC4235426. (a) 3D surface representation of ZINC4235426 (red) and BPL interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

The redocking of the co-crystallized structure of the native ligand with its protein was an essential step to confirm that the ligand bind within the binding pocket/cavity in the appropriate conformation, also the generated RMSD values of

less than 2 Å is considered for docking accuracy (Thomsen & Christensen, 2006). The redocking of native ligand (BQX) with BPL (PDB ID: 6NDL) generated the MolDock score of -177.821, and RMSD (root mean square deviation) value of 1.8 Å, which indicated that the applied protocol is favorable for docking simulation. Further, BPL is highly conserved (99.3%) in all the strains belonging to different lineages of MRSA. The multiple sequence alignment of biotin ligase protein belonging to the respective lineages of MRSA using BLOSUM62 showed the maximum score value of 3.0. In addition to this, out of 333 amino acid lengths, only two amino acid sites showed non-conservative.

5.12.7 Thymidylate kinase (TMK)

TMK is a nucleotide kinase that catalyzes deoxythymidine monophosphate to deoxythymidine diphosphate using adenosine triphosphate (ATP) as the source of the phosphoryl group and results in the biosynthesis of deoxythymidine triphosphate (dTTP) for DNA synthesis (Martínez-Botella et al., 2013). Therefore, TMK is considered an attractive potential target for antibacterial drug inhibition (Keating et al.,2012).

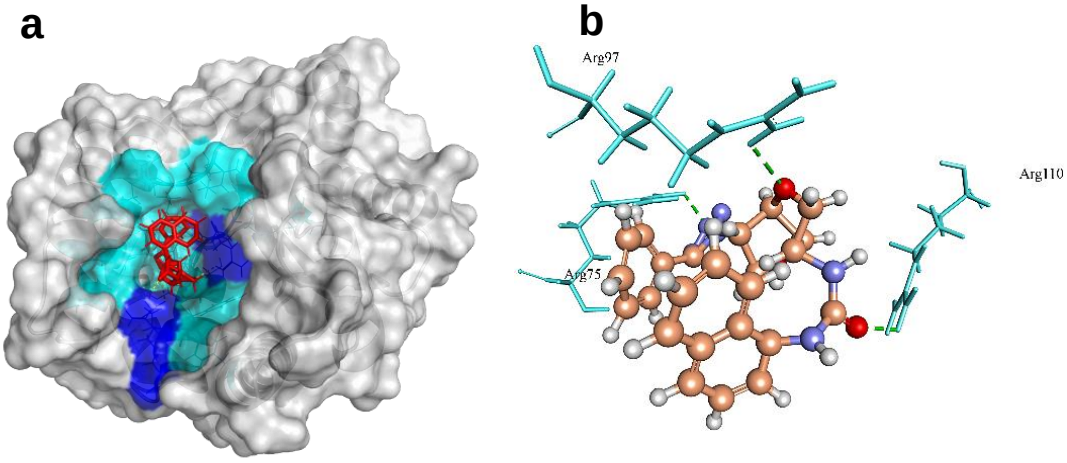


Fig. 14. The molecular docking analysis of TMK (WP_001272126.1) with compound ZINC4259578. (a) 3D surface representation of ZINC4259578 (red) and TMK interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

The crystal structure of TMK protein (4HLC from *S. aureus* subsp. *aureus* MRSA252) has active site residues (Gln101, Arg70, Arg48, and Ser97) that are

involved in H-bond interactions with native ligand benzoic acid (T05). Although none of these residues are predicted to form H-bonds with the ZINC4259578 compound. However, this compound was predicted to make 3 H-bonds with other active site residues, Arg75, Arg97, and Arg110 of TMK protein, and exhibited a MolDock score of -139.656 **(Table 3, Fig. 14)**. Also, the ZINC4235426 compound created 6 H-bonds with the active site residues, Arg75, Arg97, Glu106, Tyr105, and Glu42 of TMK protein and predicted a MolDock score of -139.150 **(Table 3)**. The redocking of native ligand (T05) with TMK (PDB ID: 4HLC) generated the MolDock score of -152.823, and RMSD value of 1.01 Å, suggesting that the applied docking protocol was highly preferred.

5.12.8 UDP-N-acetylmuramoyl alanyl-D-glutamate-2,6-diaminopimelate ligase (MurE)

The target enzyme MurE ligase is a complex molecule involved in the pathway peptidoglycan biosynthesis (Amera et al., 2020). MurE initiates reaction by adding meso-diaminopimelic acid to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanyl-d-glutamate, during the synthesis of murein in the cytoplasm (Gordon et al., 2001). This enzyme is crucial for *S. aureus* strains; therefore, it can be used as a potential antibacterial drug target (Gordon et al., 2001). Based on the interaction between the crystallographic structure of the MurE template (4C12 from *S. aureus*) and the crystallographic ligand uridine 5' diphospho N-acetyl muramoyl-L-Alanyl-D-Glutamyl-L-Lysine (UML), it was found that the active site residues involved in H-bond interactions were Ser456, Glu460, Asp406, Thr152, Ser179, Arg187, Arg383, His205, Asn151, Thr153, Thr45, Thr46, Val47, Thr28, Ser30, and Val47. However, the ZINC4235426 compound created 5 H-bonds with the active site residues, Tyr45, Thr46, Val47, and Glu155 of MurE ligase with a MolDock score of -125.654 **(Table 3, Fig. 15)**. The redocking of the native ligand (UML) with MurE (PDB ID: 4C12) generated the MolDock score of -167.754 and RMSD value of 4.75 Å. The RMSD value of redocking found above the cut-off value (<2 Å) could be the reason for the large ligand size (120 atoms), and more rotatable bond (26 flexible torsions). This finding was supported by the earlier study, which suggested that redocking accuracy decreases as the number of rotatable bonds increase regardless of the docking program used (Chen et al., 2005). Also, the typical 2 Å RMSD cut-off for docking accuracy

may not be reliable for the ligands with a large size and number of rotatable bonds (Boittier et al., 2020).

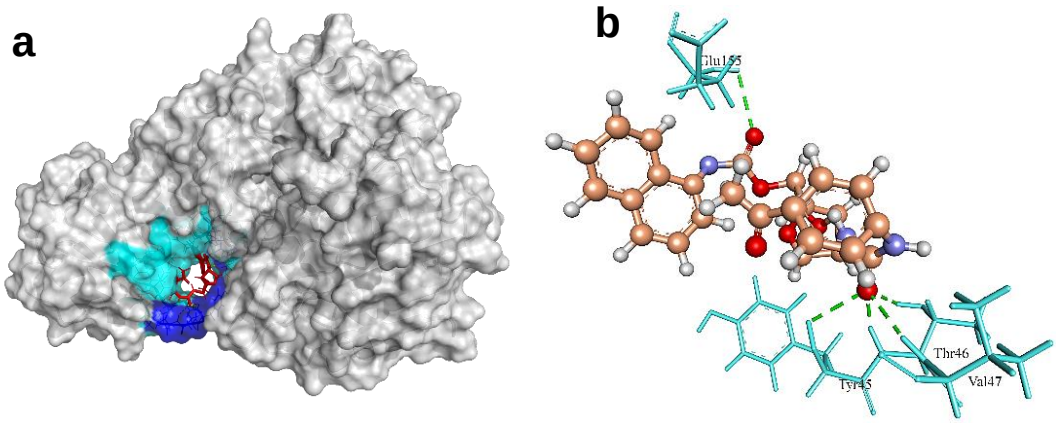


Fig. 15. The molecular docking analysis of MurE (WP_000340119.1) with compound ZINC4235426. (a) 3D surface representation of ZINC4235426 (red) and MurE interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

5.12.9 Putative phosphate acetyltransferase (Pta)

Phosphate acetyltransferase (Pta) plays an important role in acetate metabolism along with acetate kinase. This enzyme catalyzes the uptake of carbohydrates and their conversion into their respective phosphoesters during transport (Morya et al., 2012). It is also involved in other metabolic pathways such as taurine and hypotaurine metabolism, pyruvate metabolism, and propanoate metabolism (Campos-Bermudez et al., 2010). It was reported that this enzyme activity is important for virulence in pathogenic bacteria including *S. saprophyticus* (Sakinç et al., 2009). This enzyme is essential for the survival of bacteria and could be a putative drug target for the design and evaluation of a new class of antimicrobials (Morya et al., 2012). In the template crystal structure of Pta protein (4E4R from *S. aureus* subsp. *aureus* MRSA252), the native crystallographic ligand 2-amino-2-hydroxymethyl-propane-1,3-diol (TRS) created H-bonds in the active site residues are Gly130 and Asp305. However, none of these residues were involved in the H-bond formation with ZINC4270981 compound. Instead, this compound makes 2 H-bonds with the other active site residues, Leu299, and Gln325 of Pta protein and predicted the MolDock score of -134.847 **(Table 3, Fig. 16)**. The redocking of native ligand (TRS) with Pta (PDB ID: 4E4R) revealed the

MolDock score of -38.370, and RMSD value of -1.12 Å, indicating that the applied protocol was fair for this protein.

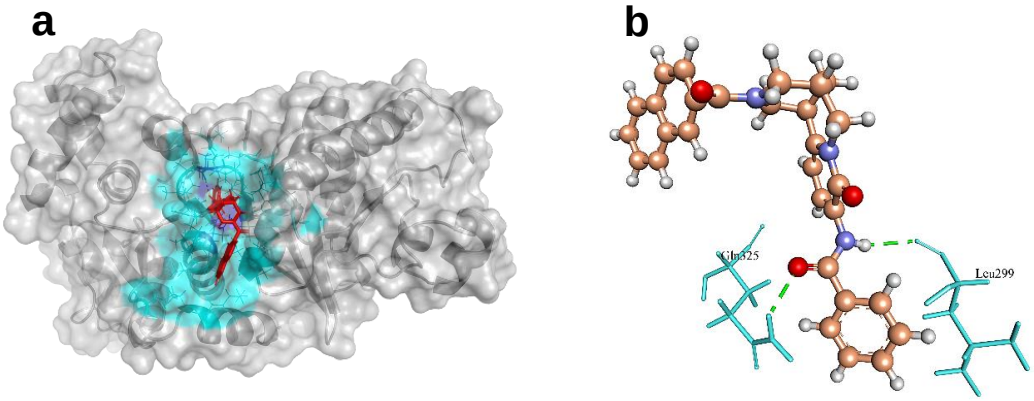


Fig. 16. The molecular docking analysis of Pta (WP_000774281.1) with compound ZINC4270981. (a) 3D surface representation of ZINC4270981 (red) and Pta interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

5.12.10 HPr kinase/phosphorylase

HPr kinase/phosphorylase is a bifunctional enzyme that enhances the glycolytic intermediates (carbon metabolism) and virulence progression in Gram-positive bacteria (Deutscher et al., 2005; Nessler, 2005). In *Listeria monocytogenes*, the metabolism of carbon sources inhibits the PrfA, transcription activator, and involves in the virulence gene expression regulation (Deutscher et al., 2005). This enzyme is of clinical interest due to its regulatory roles in the infectious process and therefore could be a new drug target. The template crystal structure of HPr kinase/P (1KO7 from *S. xylosus*) interacted with native crystallographic ligand PO_4. The cavity 1 (S:69.512 Å; V:213.76 Å; R:18 Å) of Hpr Kinase/P defined by MVD tool was selected for docking. The ZINC4235426 compound formed 4 H-bonds with the active site residues, Lys259, Thr150, and Asn227, and generated the MolDock score of -147.451 **(Table 3, Fig. 17a & 17b)**. Also, the ZINC4235924 compound formed 7 H-bonds with the active site residues, Gly151, Thr150, Asn227, Lys258, Thr260, and Asn229 of HPr kinase/P and revealed the MolDock score of -137.549 **(Table 3, Fig. 17c & 17d)**.

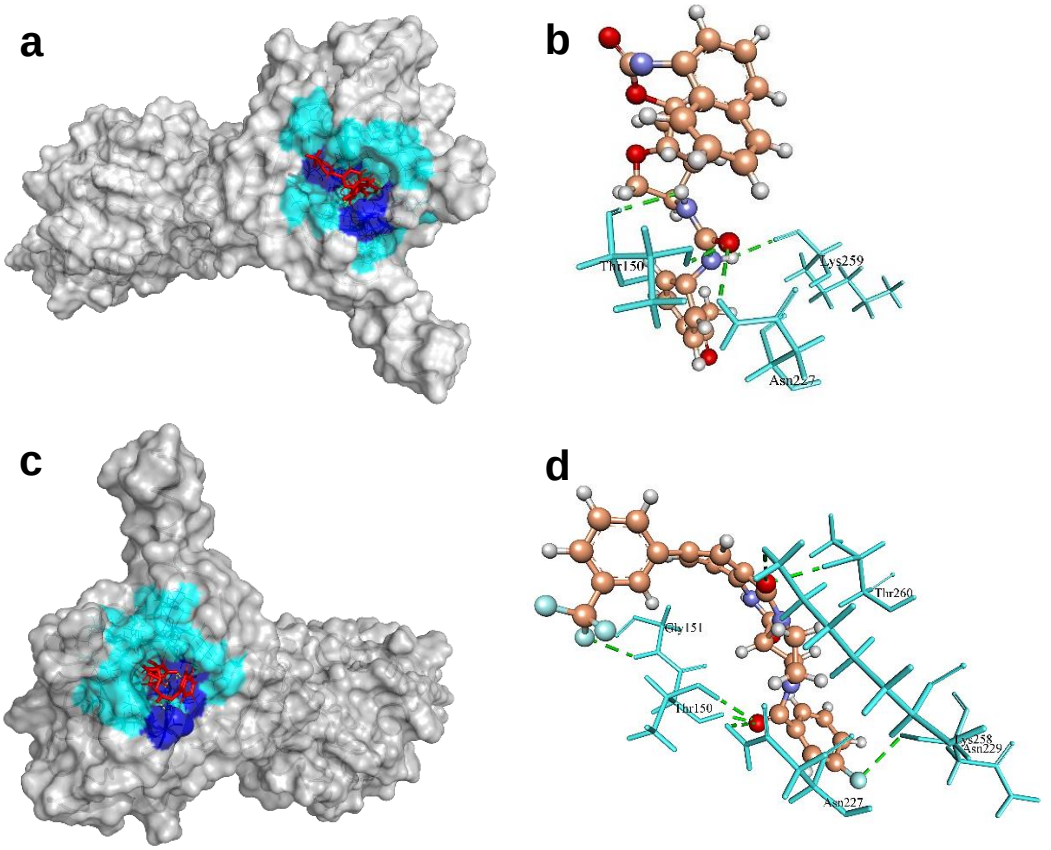


Fig. 17. The molecular docking analysis of HPr kinase (WP_000958224.1) with compounds ZINC4235426, and ZINC4235924, respectively. (a & c) 3D surface representation of ZINC4235426 (red), ZINC4235924 (red), and HPr kinase interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b & d)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

5.12.11UTP-glucose-1-phosphate uridylyltransferase (UGPase)

This enzyme is also known as UDP-glucose pyrophosphorylase or UGPase, is ubiquitous due to its important role in glycogen synthesis and production of glycolipids, glycoproteins, and proteoglycans (Berbis et al., 2015; Thoden & Holden, 2007). It is also required for the capsular polysaccharide biosynthesis and plays an essential role as a virulence factor in *Streptococcus pneumoniae* (Bonofiglio et al., 2005). A defective UGPase fails to incorporate galactose into its cell wall (Bonofiglio et al., 2005). The indispensability of the UGPase projects the enzyme as a potential drug target (Genevaux et al., 1999). In the template crystal structure of UGPase (5VCT from *Burkholderia ambifaria* MC40-6), the native crystallographic ligand citric acid (CIT) created 2 H-bonds in the active site residues are Lys16, and Leu14. However,

none of these residues were involved in the H-bond formation with compounds. Instead, ZINC428871 compound formed a single H-bond with the other active site residues, Leu110 of UGPase, and predicted the MolDock score of -122.664 **(Table 3, Fig. 18a & 18b)**. Also, the ZINC31154666 compound makes 4 H-bonds with the other active site residues, Gly171, Leu230, Leu110, and Asn226 of UGPase and predicted the MolDock score of -120.197 **(Table 3)**. The redocking of native ligand (CIT) with UGPase (PDB ID: 5VCT) achieved MolDock score of 29.116, and RMSD value of 1.84, suggesting that the applied docking simulation protocol was satisfied for this protein.

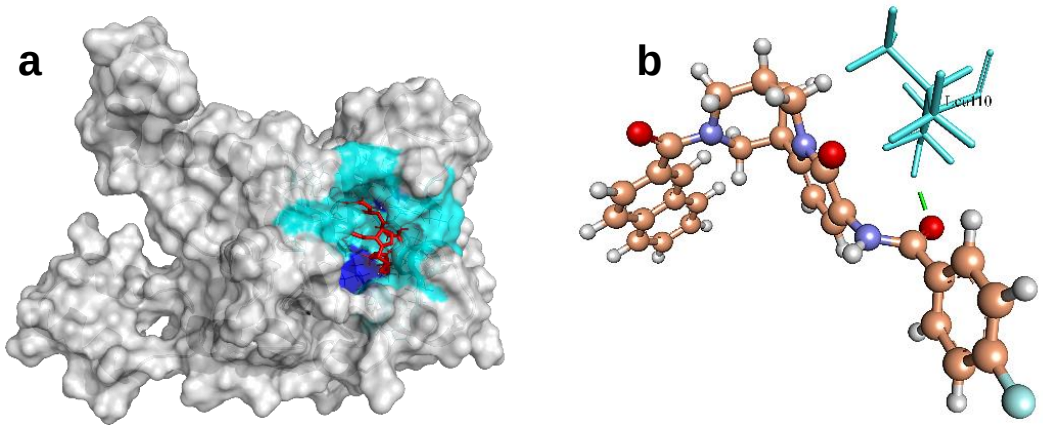


Fig. 18. The molecular docking analysis of UGPase (WP_000721337.1) with compounds ZINC428871. (a) 3D surface representation of ZINC428871 (red) and MurE interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

5.12.12 Putative fatty acid synthesis protein (PlsX)

Putative fatty acid synthesis protein (PlsX) is the key enzyme that coordinates the fatty acid synthase II (FASII) pathway to the phospholipid synthesis pathway and plays an essential passage of unsaturated fatty acids into the membrane (Kim et al., 2009). FASII is the process used by bacteria to produce the fatty acid components of phospholipids which is essential in human pathogens (Lu et al., 2006). The essential role of fatty acid in membrane structure has provided attention to targeting this pathway (Yao & Rock, 2018). The B chain structure of 1U7N template from *Enterococcus faecalis* for PlsX has no native ligand. Therefore, cavity 1 (V:180.224 Å; S:660.48 Å, R:19 Å) identified by MVD tool was selected for docking with

ZINC4237105 compound. This compound formed 1 H-bond to the active site residue, Lys262 of PlsX protein, and generated MolDock score of -130.756 **(Table 3, Fig. 19)**.

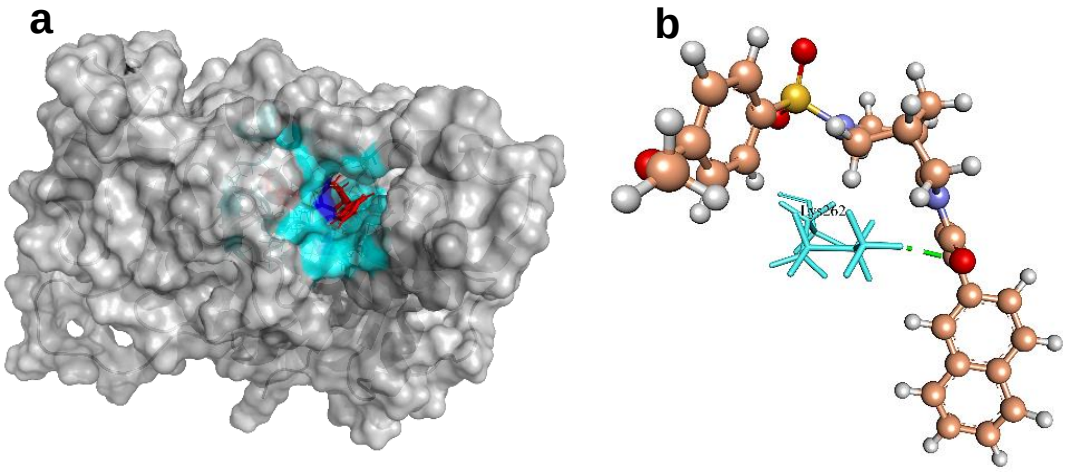


Fig. 19. The molecular docking analysis of PlsX (WP_000239744.1) with compound ZINC4237105. (a) 3D surface representation of ZINC4237105 (red) and PlsX interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residue (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

5.12.13 Pantoate beta alanine ligase (PanC)

Pantoate beta-alanine ligase (PanC) is the last enzyme involved in pantothenate biosynthesis. It catalyzes the adenosine triphosphate (ATP)-dependent condensation of pantoate and β-alanine to form pantothenate (vitamin B5) (von Delft et al., 2001). It is essential for the growth and survival of *S. aureus* by playing a critical role in fatty acid metabolism, identified as a potential target for new antimicrobials (Pradhan & Sinha, 2018). Based on a structural comparison with a crystallographic structure of PanC template (PDB ID: 3AG6 from *S. aureus* subsp. *aureus* NCTC 8325), the active site residues involved in H-bonds with the crystallographic native ligand pantoyl adenylate (PAJ) were Gln154, Gln62, Met31, Gly148, His35, His38, Lys185, and Val177. The ZINC4235426 compound created four H-bonds with the active site residues, Met31, Gly148, His35, and Thr30 of PanC protein, with a MolDock score of -173.843 **(Table 3, Fig. 20)**. The redocking of native ligand (PAJ) with PanC (PDB ID: 3AG6) generated the MolDock score of -174.601, and the RMSD value of -1.57 Å, indicating that the used protocol for docking simulation was highly preferable.

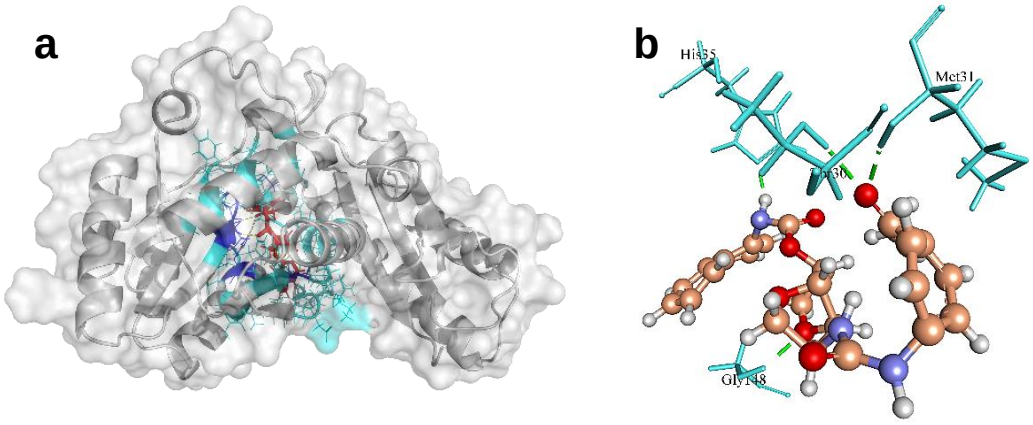


Fig. 20. The molecular docking analysis of PanC (WP_000163742.1) with compound ZINC4235426. (a) 3D surface representation of ZINC4235426 (red) and PanC interactions with hydrogen bonding sites (blue), and hydrophobic interactions (cyan). **(b)** Residues (cyan) involved in the H-bond interaction (green dashed lines) with the compound (scaled ball and stick).

The SwissADME analysis results showed that all the drug molecules listed in **Table 2** satisfy Lipinski's rule of five with zero violations and had no PAIN alerts. Additionally, ZINC4235426, ZINC4259578, and ZINC428871 assured the properties of drug-likeness (Ghose, Veber, Egan, and Muegge). The *K*p values of the drug molecules are in the ranges of -6.21 to -7.32 cm/s suggesting low skin permeability **(**Daina et al., 2017**)**. The ZINC4235924 and ZINC4270981 molecules were found to have high blood-brain barrier (BBB) permeability. All the compounds are substrates of permeability glycoprotein (P-gp) except ZINC428871. The logP values are predicted in the ranges of 0.82 to 4 indicating that the drug molecules have optimal lipophilicity. The drug molecules showed high gastrointestinal (GI) absorption and bioavailability scores, and have non-carcinogenicity.

5.13 Potential vaccine target candidates

In antigen-based vaccine design, adhesins are key molecules for the design of vaccines against microbial infective agents (Sachdeva et al., 2005). With the advancement of bioinformatics, screening of adhesion molecules in the bacterial genome has become quicker through homology-based approaches. SPAAN (software program for the prediction of adhesins and adhesin-like proteins using a neural network) program integrated with Vaxign v.2.0 tool is based on highly curated datasets and neural networks that were optimally trained for compositional attributes.

The probability for a given protein likely to be an adhesin is the weighted average of individual probabilities emerging from the five networks, based on the accuracy of each network (Vivona et al., 2006). The search for candidate adhesin proteins can be a most promising approach for identifying novel vaccine candidates (Solanki et al., 2019).

The previous study on MRSA 252 genome revealed that IsaA, HlgA, SsaA, and IsdB were the best immunogenic targets (Noori et al., 2021). In the present study, the 20 proteins were localized in secreted (SEC), surface-exposed (PSE), and membrane proteins (MEM) and whose structures were evaluated for adhesion capability to MHC-I and II then found that four proteins have adhesion indices higher than 0.51, which means that they may induce either cellular or humoral adaptive immune responses (Unal & Steinert, 2014). Among these proteins, Vaxign v.2.0 identified the 4 best putative vaccine candidates *viz*., foldase protein, ESAT-6 machinery protein, penicillin-binding protein (PBP) 1, and PBP2 **(Table 4)**.

Foldase protein (PrsA) (WP_000782119.1) was found to be the most promising vaccine candidate among other candidates since this membrane protein has an accessory role in virulence (Unal & Steinert, 2014), high antigenicity (0.7662), adherence score (0.68), low molecular weight (35.623 kDa), no transmembrane helix, and has lipoprotein signal peptide at 20-21 amino acid sequence position **(Table 4)**. PrsA has a PPIC-type PPIASE domain belonging to the Rotamase family (PF00639), this domain region is in 146-245 aa (amino acid) positions of the protein sequence, and acts as interconversion of cis-proline and trans-proline. It was reported that PrsA is a surface-exposed protein that is essential for protein folding and involve in cell wall biosynthesis and bacterial pathogenicity (Cron et al., 2009; Jakob et al., 2015). This protein stimulates an antibody response and extends the protection against multiple mouse infections (Henningham et al., 2012; Nanduri et al., 2008). In *Streptococcus sanguinis*, this protein is revealed as the vaccine candidate with the induction of opsonic antibodies (Ge et al., 2010). Further, PrsA is highly conserved in *Legionella pneumophila* and found immunogenic, thus offering the scope for the development of a DNA vaccine (Humbert et al., 2015). This study found that PrsA is highly conserved (99.6%) in various lineages of MRSA, and suggested that PrsA from MRSA could provide protection against the different lineages of pathogenic *S. aureus* strains and may be a novel protective antigen for MRSA vaccine development.

Table 4. Vaccine target candidates for *S. aureus* identified by Vaxign *v*2.0

Vaccine Target	Length (aa)	Mol. Wt. (kDa)	Adh score	Ep. No.	VaxiJen score	TM	Sig. P
Foldase protein (PrsA)	320	35.623	0.68	5	0.7662	0	Yes, 20-21
ESAT-6 machinery protein (EssA)	152	17.392	0.58	1	0.7034	1	No
Penicillin-binding protein 1 (PBP1)	744	82.738	0.53	10	0.6351	1	No
DD-transpeptidase (PBP2)	727	80.356	0.82	10	0.6846	1	No

aa represents the amino acid; **Mol. Wt. (kDa)** indicates molecular weight in kilo Delton**; Adh score** indicates adhesion score defined by VaxiJen; **Ep. Nos.** indicate the number of B-cell epitopes present in the protein**; VaxiJen score** indicates the score for a suitable vaccine candidate; TM indicates the number of a transmembrane helix; and **Sig. P** indicates the signal peptide presence or absence in the amino-acid positions.

The ESAT-6 machinery protein (EssA) protein is the conserved membrane protein that is necessary for the synthesis and secretion of EsxA protein (Zhou et al., 2013). Secretion of EsxA was prevented in the absence of the *essA* gene and this protein might play a role in the process of the pathogenesis for *S. aureus*. Previous studies indicated that EsxA was the important candidate antigen for the *S. aureus* vaccine development (Gröschel et al., 2016). Immune protective antigen, EssA is a highly homologous protein of *Mycobacterium tuberculosis* and has good virulence and immunogenicity (Zarantonelli et al., 2006). EssA protein can induce a high level of an immune response against *Streptococcus agalactiae* infection (Ma et al., 2020). In this study, EssA protein (WP_000928935.1) of *S. aureus* was found with a T7SS_EssA_Firm domain (PF10661) located at 2-144 aa positions in its sequence. This protein was found better MHC adhesion capacity with an adhesion score of 0.58, high antigenicity (0.7034), and nine epitopes **(Table 4)**. Also, these proteins have the epitopes of both B-cell and T-cell that can induce host immune responses. Considering these criteria, our finding suggests that this protein could potentially be vaccine antigens against the pathogenic *S. aureus* strains.

Every core-genome of *S. aureus* possesses four penicillin-binding proteins (PBP1-PBP4) that are important for peptidoglycan biosynthesis (Zarantonelli et al.,

2006). Among these PBPs, PBP1 and PBP2 play important roles in *S. aureus* survival (Da Costa et al., 2018). Also, these proteins have immunogenic properties and suggested that these proteins could be vaccine candidates in *N. meningitidis* (Monterrubio-López et al., 2015). It was reported that PBP1 from *M. tuberculosis* can be used to design a new TB vaccine (Zarantonelli et al., 2006). Another experiment suggested that vaccination with PBP2 induces protection against a protein that is involved in chromosome-mediated antibiotic resistance in meningococcal disease and proposed that this protein could be a promising vaccine candidate (Barh et al., 2011; Pizza et al., 2000). The information on these two proteins for the vaccine development against *S. aureus* is scantly available. In this study, it was found that PBP1 and PBP2 proteins have the best characteristics of antigenic value, MHC good adhesion capacity, stability, and epitope numbers **(Table 4)**. PBP1 protein has three conserved domains, FtsI (COG0768), PASTA_PBP2x-like_1 (cd06576), and PASTA_PBP2x-like_2 (cd06575). FtsI domain is found in 28-592 aa regions of the protein sequence, controlling cell cycle, cell division, chromosome partitioning, and cell wall or membrane biogenesis. PASTA domains of PBP2x-like_1, and PBP2x-like_2 are found in the 601-655, and 659-712 aa regions of PBP1 protein sequence. These domains catalyze the peptidoglycan synthesis which is essential for cell division and protects from osmotic shock and lysis. PBP2 is a DD-transpeptidase essential for bacterial cell wall synthesis. This protein has a conserved domain - MrcB (membrane carboxypeptidase B), which is found in the 46-727 aa regions of the sequence which bind to the β-lactam thiazolidine ring system of β-lactam antibiotics.

The identified vaccine candidate proteins are localized in MEM regions and have high stability, antigenic and non-allergen properties **(Table 4)**. Also, these proteins have an adhesion score > 0.51, ensuring that they can efficiently bind to MHC class I and II molecules, and may induce either cellular or humoral adaptive immune responses (Unal & Steinert, 2014). It was reported that clinical trials to design effective anti-Staphylococcal vaccines have not been fruitful (Tahir ul Qamar et al., 2021), thus numerous vaccines were reported that mainly focus to trigger B-cell response and development of antibodies opsonization (Soltan et al., 2020). Additionally, the presence of antigenic B-cell epitopes in these candidate proteins confirms the ability to interact with the MHC class I molecule. The interaction of MHC class I presented antigens with cytotoxic CD8+ lymphocytes is one of the potential vaccine-induced immune responses (Kim et al., 2010). The prospective

candidate proteins reported in this study fulfill all the prerequisites (subcellular localization, antigenic and adhesin properties) of being potent vaccine candidates. Among the identified vaccine candidates, PrsA and EssA proteins have higher antigenic properties defined by Vaxijen than PBP1 and PBP2 proteins **(Table 4)**. Previous studies suggested that the protein with the highest antigenicity could be recognized by immune response easily and evoke the immune responses (Jiang et al., 2019; Monterrubio-López et al., 2015; Unal & Steinert, 2014; Zhou et al., 2013); therefore, PrsA and EssA proteins could be effective vaccine candidates. Further, the multiple epitopes from the identified vaccine candidates having different pathways could be used to develop a universal anti-staphylococcal vaccine (Solanki et al., 2021).

The proposed potential vaccine candidates selected in this report have antigenic peptides with the ability to induce the humoral and cell-mediated immune responses, thus these vaccine candidate proteins could be applied in several vaccine designs to perform web-lab experiments to validate them as DNA vaccines or recombinant proteins, to develop protection against staphylococcal infections.

6. Conclusion

MRSA is the leading cause of nosocomial and community infections and the emergence of hypervirulent strains and becoming a greater threat to the public. The phenotypic and genotypic characterizations are important for identifying the risk factors associated with *S. aureus* infections and are useful to monitor and control the circulation or transmission of these strains. Furthermore, the development of alternatives to antibiotics for the treatment and prevention of staphylococcal infections is of great concern. Genomic-wide comparative analysis of such pathogens could extend our understanding of pathogenesis and evolution at the molecular level and has the potential to a breakthrough in diagnosis, treatment, and infection control.

In this study, the characterization of MRSA through a polyphasic approach could determine the relatedness among geographically diverse MRSA strains, however, the data provided by such an approach could not provide better knowledge to understand the molecular basis of drug-resistance, pathogenesis, niche-specific difference, and evolutionary relationship of closely related *S. aureus* clinical strains. The data generated from the WGS confirmed the diversity of MRSA strains among the same CC5, CC8, CC45, and CC22. The comparative genome analysis provides high resolution to distinguish between the closely related sequenced strains which are

indistinguishable by a polyphasic approach. Also, it allowed the segregation of isolates of geographical origin and differentiation of clinical isolates from the commensal isolates. The genome mining approach has clearly stated that the biofilm-forming ability of MRSA was not correlated with the presence of biofilm-forming encoding genes, also the genetic constituents have no information regarding the infection sites. An interesting finding is the addition of the SA G6 genome responsible for open pan-genome and diversity among genomes. The openness of pan-genomes of *S. aureus* isolates relies on the acquisition of MGEs. The evidence of MGEs transfer event especially in SA G6 is supported by the drastic drop of the core/pan-genome ratio curve, and gaps and GC skewed regions in the comparative genome map. The presence of *ant(6)-Ia, aph(30)-III)* and *sat-4* in the GI region of SA G6 are likely acquired and these genes may provide fitness and a selective advantage during host-adaptation and colonization. The core-genome and accessory genome revealed that *S. aureus* isolates required amino acids than carbohydrates as the energy source and suggests that these isolates adapted to grow in a protein-rich medium than carbohydrates. Phylogenetic analysis suggests that SA G6 and *S. aureus* subsp. *aureus* ST228 strains are distinct from their group. The acquisition of plasmid, prophage functional modules, ARGs, and VFGs in *S. aureus* isolates contributes a major role in the rapid evolution of pathogenic *S. aureus* lineages and that confer specific advantages in a defined host under environmental conditions. This comparative genome analysis would improve the knowledge about the pathogenic *S. aureus* strain's characterization, adaptation, and dynamic evolutionary process in the transmission of infections globally. The mining of prophage signal from *S. aureus* genomic data suggested that the frequency of phage-mediated HGT is higher between the *S. aureus* strains with the same *SCCmec* type, STs, or CCs. The presence of similar genetic elements in the prophages isolated from *S. aureus* associated with animals and humans suggested that prophages may have played a major role in the epidemiological changes. The appearance of the mosaic nature of prophage genomes suggested the occurrence of genetic exchange among the *S. aureus* strains *via* phages. The IEC is highly human-specific, however, this computational study finding revealed that the presence of IEC could not differentiate between phages of human and animal-associated *S. aureus.* The acquisition of various VFGs in prophages of animal-associated *S. aureus* suggested that these prophages could be more pathogenic potential than prophages of human-associated *S. aureus.* Remarkably, this study

showed evidence of phage-mediated HGT between the *S. aureus* associated with bovine (cow) infections and the *S. aureus* associated with dog infections. The prophages carried by human-associated *S. aureus* strains with different serotypes and obtained from different geographical locations scattered in all the clades, suggesting that these phages have a wide distribution across the European regions and among the various *S. aureus* serotypes. Comparative studies of prophages carried by human and animal-associated *S. aureus* strains are of very crucial importance for the investigation of *S. aureus* transmission from human to animal and *vice-versa*.

In the present study, subtractive genome and reverse vaccinology approaches were applied to predict potential drug/vaccine molecules which can be used in the development of promising drugs and vaccines for MDR *S. aureus.* These in-silico approaches have identified eight potential drug targets for MRSA **(Table 3)**. The molecule docking analysis result suggested that the ZINC4235426 compound is a promising drug molecule. Also, using the reverse vaccinology approach, 4 putative antigenic proteins were identified, among these, PrsA and EssA proteins were found to be more promising vaccine candidates **(Table 4)**. These identified proteins can be used for further rational drug or vaccine design to identify novel therapeutic agents for the treatment of MDR staphylococcal infections. Further, the identified targets through in-silico approaches are required to perform validation using *in-vitro* and *in-vivo* experiments to find new methods of treating the diseases caused by MRSA strains.

7. References

Aanensen, D. M., Feil, E. J., Holden, M. T. G., Dordel, J., Yeats, C. A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., Chlebowicz, M. A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Åhman, J., Matuschek, E., Friedrich, A. W., … Kearns, A. (2016). Whole-genome sequencing for routine pathogen surveillance in public health: A population snapshot of invasive Staphylococcus aureus in Europe. *MBio*, *7*(3). https://doi.org/10.1128/mBio.00444-16

Abadio, A. K. R., Kioshima, E. S., Teixeira, M. M., Martins, N. F., Maigret, B., & Felipe, M. S. S. (2011). Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-75

Abi Hussein, H., Geneix, C., Petitjean, M., Borrel, A., Flatters, D., & Camproux, A. C. (2017). Global vision of druggability issues: applications and perspectives. In *Drug Discovery Today* (Vol. 22, Issue 2, pp. 404–415). Elsevier Ltd. https://doi.org/10.1016/j.drudis.2016.11.021

Aires De Sousa, M., & de Lencastre, H. (2004). Bridges from hospitals to the laboratory: Genetic portraits of methicillin-resistant Staphylococcus aureus clones. In *FEMS Immunology and Medical Microbiology* (Vol. 40, Issue 2, pp. 101–111). Elsevier. https://doi.org/10.1016/S0928-8244(03)00370-5

Aires-de-Sousa, M. (2017). Methicillin-resistant Staphylococcus aureus among animals: current overview. In *Clinical Microbiology and Infection* (Vol. 23, Issue 6, pp. 373–380). Elsevier B.V. https://doi.org/10.1016/j.cmi.2016.11.002

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. v., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., … McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517–D525. https://doi.org/10.1093/nar/gkz935

Al-Obaidi, M. M. J., Suhaili, Z., & Desa, M. N. M. (2018). Genotyping Approaches for Identification and Characterization of Staphylococcus aureus. In *Genotyping*. https://doi.org/10.5772/intechopen.75969

Al-Tam, F., Brunel, A. S., Bouzinbi, N., Corne, P., Bañuls, A. L., & Shahbazkia, H. R. (2012). DNAGear- A free software for spa type identification in Staphylococcus aureus. *BMC Research Notes*, *5*. https://doi.org/10.1186/1756-0500-5-642

Amera, G. M., Khan, R. J., Jha, R. K., Pathak, A., Muthukumaran, J., & Singh, A. K. (2020). Prioritization of Mur family drug targets against A. baumannii and identification of their homologous proteins through molecular phylogeny, primary sequence, and structural analysis. *Journal of Genetic Engineering and Biotechnology*, *18*(1), 1–22. https://doi.org/10.1186/s43141-020-00048-4

Amineni, U., Pradhan, D., & Marisetty, H. (2010). In silico identification of common putative drug targets in Leptospira interrogans. *Journal of Chemical Biology*, *3*(4), 165–173. https://doi.org/10.1007/s12154-010-0039-1

Appelbaum, P. C. (2007). Microbiology of antibiotic resistance in Staphylococcus aureus. In *Clinical Infectious Diseases* (Vol. 45, Issue SUPPL. 3). Clin Infect Dis. https://doi.org/10.1086/519474

Archer, N. K., Mazaitis, M. J., William Costerton, J., Leid, J. G., Powers, M. E., & Shirtliff, M. E. (2011). Staphylococcus aureus biofilms: Properties, regulation and roles in human disease. In *Virulence* (Vol. 2, Issue 5, pp. 445–459). Taylor and Francis Inc. https://doi.org/10.4161/viru.2.5.17724

Arciola, C. R., Baldassarri, L., & Montanaro, L. (2001). Presence of icaA and icaD genes and slime production in a collection of Staphylococcal strains from catheter-associated infections. *Journal of Clinical Microbiology*, *39*(6), 2151–2156. https://doi.org/10.1128/JCM.39.6.2151-2156.2001

Arciola, C. R., Campoccia, D., Baldassarri, L., Donati, M. E., Pirini, V., Gamberini, S., & Montanaro, L. (2006). Detection of biofilm formation in Staphylococcus epidermidis from implant infections. Comparison of a PCR-method that recognizes the presence of ica genes with two classic phenotypic methods. *Journal of Biomedical Materials Research - Part A*, *76*(2), 425–430. https://doi.org/10.1002/jbm.a.30552

Arciola, C. R., Campoccia, D., Ravaioli, S., & Montanaro, L. (2015). Polysaccharide intercellular adhesin in biofilm: Structural and regulatory aspects. In *Frontiers in Cellular and Infection Microbiology* (Vol. 5, Issue FEB). Frontiers Research Foundation. https://doi.org/10.3389/fcimb.2015.00007

Argudín, M. Á., Mendoza, M. C., & Rodicio, M. R. (2010). Food Poisoning and Staphylococcus aureus Enterotoxins. In *Toxins* (Vol. 2, Issue 7). https://doi.org/10.3390/toxins2071751

Armand-Lefevre, L., Ruimy, R., & Andremont, A. (2005). Clonal comparison of Staphylococcus from healthy pig farmers, human controls, and pigs. *Emerging Infectious Diseases*, *11*(5), 711–714. https://doi.org/10.3201/eid1105.040866

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Asadollahi, P., Farahani, N. N., Mirzaii, M., Khoramrooz, S. S., van Belkum, A., Asadollahi, K., Dadashi, M., & Darban-Sarokhalil, D. (2018). Distribution of the Most Prevalent Spa Types among Clinical Isolates of Methicillin-Resistant and -Susceptible Staphylococcus aureus around the World: A Review. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00163

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*. https://doi.org/10.1186/1471-2164-9-75

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K. I., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., & Hiramatsu, K. (2002).

Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*, *359*(9320), 1819–1827. https://doi.org/10.1016/S0140-6736(02)08713-5

Barh, D., Tiwari, S., Jain, N., Ali, A., Santos, A. R., Misra, A. N., Azevedo, V., & Kumar, A. (2011). In Silico Subtractive Genomics for Target Identification in Human Bacterial Pathogens. *Research Overview Drug Dev Res*, *72*, 162–177. https://doi.org/10.1002/ddr.20413

Barrera-Rivas, C. I., Valle-Hurtado, N. A., González-Lugo, G. M., Baizabal-Aguirre, V. M., Bravo-Patiño, A., Cajero-Juárez, M., & Valdez-Alarcón, J. J. (2017). Bacteriophage Therapy: An Alternative for the Treatment of Staphylococcus aureus Infections in Animals and Animal Models. In *Frontiers in Staphylococcus aureus*. InTech. https://doi.org/10.5772/65761

Barrett, L., & Atkins, B. (2014). The clinical presentation of prosthetic joint infection. *Journal of Antimicrobial Chemotherapy*, *69*(SUPPL1). https://doi.org/10.1093/jac/dku250

Bartlett, A. H., & Hulten, K. G. (2010). Staphylococcus aureus pathogenesis: Secretion systems, adhesins, and invasins. In *Pediatric Infectious Disease Journal* (Vol. 29, Issue 9, pp. 860–861). Lippincott Williams and Wilkins. https://doi.org/10.1097/INF.0b013e3181ef2477

Baumler, A., & Fang, F. C. (2013). Host Specificity of Bacterial Pathogens. *Cold Spring Harbor Perspectives in Medicine*, *3*(12), a010041–a010041. https://doi.org/10.1101/cshperspect.a010041

ben Zakour, N. L., Guinane, C. M., & Fitzgerald, J. R. (2008). Pathogenomics of the staphylococci: Insights into niche adaptation and the emergence of new virulent strains. In *FEMS Microbiology Letters* (Vol. 289, Issue 1, pp. 1–12). FEMS Microbiol Lett. https://doi.org/10.1111/j.1574-6968.2008.01384.x

Berbis, M., Sanchez-Puelles, J., Canada, F., & Jimenez-Barbero, J. (2015). Structure and Function of Prokaryotic UDP-Glucose Pyrophosphorylase, A Drug Target Candidate. *Current Medicinal Chemistry*, *22*(14). https://doi.org/10.2174/0929867322666150114151248

Bergey, D. H., & Holt, J. G. (1994). *Bergey's Manual of Determinative Bacteriology* (9th ed.). Williams & Wilkins.

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., & Brinkman, F. S. L. (2017). IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, *45*(W1), W30–W35. https://doi.org/10.1093/nar/gkx343

Betley, M. J., & Mekalanos, J. J. (1985). Staphylococcal enterotoxin a is encoded by phage. *Science*, *229*(4709), 185–187. https://doi.org/10.1126/science.3160112

Bhakdi, S., & Tranum-Jensen, J. (1991). Alpha-toxin of Staphylococcus aureus. In *Microbiological Reviews* (Vol. 55, Issue 4, pp. 733–751). American Society for Microbiology (ASM). https://doi.org/10.1128/mmbr.55.4.733-751.1991

Bhakdi, S., Walev, I., Husmann, M., & Valeva, A. (2004). *Staphylococcal alpha-toxin* (pp. 91–110). Springer, Berlin, Heidelberg. https://doi.org/10.1007/b100513

Bhutia, K. O., Singh, T. S. K., Adhikari, L., & Biswas, S. (2015). Molecular characterization of community- & hospital-acquired methicillin-resistant & methicillin-sensitive Staphylococcus aureus isolates in Sikkim. *Indian Journal of Medical Research*, *142*(September), 330–335. https://doi.org/10.4103/0971-5916.166600

Bikandi, J., Millán, R. S., Rementeria, A., & Garaizar, J. (2004). In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics*, *20*(5), 798–799. https://doi.org/10.1093/bioinformatics/btg491

Bitencourt-Ferreira, G., & de Azevedo, W. F. (2019). Molegro virtual docker for docking. In *Methods in Molecular Biology* (Vol. 2053, pp. 149–167). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9752-7_10

Bitrus, A. A., Zunita, Z., Bejo, S. K., Othman, S., & Nadzir, N. A. A. (2017). In vitro transfer of methicillin resistance determinants mecA from methicillin resistant Staphylococcus aureus (MRSA) to methicillin susceptible Staphylococcus aureus (MSSA). *BMC Microbiology*, *17*(1). https://doi.org/10.1186/s12866-017-0994-6

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1). https://doi.org/10.1093/nar/gkz310

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., & Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Research*, *44*(W1). https://doi.org/10.1093/nar/gkw255

Boakes, E., Kearns, A. M., Ganner, M., Perry, C., Hill, R. L., & Ellington, M. J. (2011). Distinct Bacteriophages Encoding Panton-Valentine Leukocidin (PVL) among International Methicillin-Resistant Staphylococcus aureus Clones Harboring PVL. *Journal of Clinical Microbiology*, *49*(2), 684. https://doi.org/10.1128/JCM.01917-10

Boittier, E., Tang, Y., Buckley, M., Schuurs, Z., Richard, D. and Gandhi, N., 2020. Assessing Molecular Docking Tools to Guide Targeted Drug Discovery of CD38 Inhibitors. International Journal of Molecular Sciences, 21(15), p.5183

Boles, B. R., & Horswill, A. R. (2008). agr-mediated dispersal of Staphylococcus aureus biofilms. *PLoS Pathogens*, *4*(4). https://doi.org/10.1371/journal.ppat.1000052

Boles, B. R., Thoende, M., Roth, A. J., & Horswill, A. R. (2010). Identification of genes involved in polysaccharide- independent Staphylococcus aureus biofilm formation. *PLoS ONE*, *5*(4). https://doi.org/10.1371/journal.pone.0010146

Bonofiglio, L., García, E., & Mollerach, M. (2005). Biochemical Characterization of the Pneumococcal Glucose 1-Phosphate Uridylyltransferase (GalU) Essential for Capsule Biosynthesis. *Current Microbiology*, *51*(4). https://doi.org/10.1007/s00284-005-4466-0

Bos, K. I., Kühnert, D., Herbig, A., Esquivel-Gomez, L. R., Andrades Valtueña, A., Barquera, R., Giffin, K., Kumar Lankapalli, A., Nelson, E. A., Sabin, S., Spyrou, M. A., & Krause, J. (2019). Paleomicrobiology: Diagnosis and evolution of ancient pathogens. In *Annual*

Review of Microbiology (Vol. 73). https://doi.org/10.1146/annurev-micro-090817-062436

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., Crescenzi, P., Fani, R., & Fondi, M. (2015). MeDuSa: A multi-draft based scaffolder. *Bioinformatics*, *31*(15). https://doi.org/10.1093/bioinformatics/btv171

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. (2016). Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), E3801–E3809. https://doi.org/10.1073/pnas.1523199113

Boss, R., Cosandey, A., Luini, M., Artursson, K., Bardiau, M., Breitenwieser, F., Hehenberger, E., Lam, T., Mansfeld, M., Michel, A., Mösslacher, G., Naskova, J., Nelson, S., Podpečan, O., Raemy, A., Ryan, E., Salat, O., Zangerl, P., Steiner, A., & Graber, H. U. (2016). Bovine Staphylococcus aureus: Subtyping, evolution, and zoonotic transfer. *Journal of Dairy Science*, *99*(1), 515–528. https://doi.org/10.3168/jds.2015-9589

Botka, T., Růžičková, V., Konečná, H., Pantůček, R., Rychlík, I., Zdráhal, Z., Petráš, P., & Doškař, J. (2015). Complete genome analysis of two new bacteriophages isolated from impetigo strains of Staphylococcus aureus. *Virus Genes*, *51*(1). https://doi.org/10.1007/s11262-015-1223-8

Bukowski, M., Wladyka, B., & Dubin, G. (2010). Exfoliative toxins of Staphylococcus aureus. In *Toxins* (Vol. 2, Issue 5). https://doi.org/10.3390/toxins2051148

Burts, M. L., Williams, W. A., DeBord, K., & Missiakas, D. M. (2005). EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4). https://doi.org/10.1073/pnas.0405620102

Cafiso, V., Bertuccio, T., Santagati, M., Demelio, V., Spina, D., Nicoletti, G., & Stefani, S. (2007). agr-Genotyping and transcriptional analysis of biofilm-producing Staphylococcus aureus. *FEMS Immunology and Medical Microbiology*, *51*(1). https://doi.org/10.1111/j.1574-695X.2007.00298.x

Cameron, D. R., Jiang, J. H., Kostoulias, X., Foxwell, D. J., & Peleg, A. Y. (2016). Vancomycin susceptibility in methicillin-resistant Staphylococcus aureus is mediated by YycHI activation of the WalRK essential two-component regulatory system. *Scientific Reports*, *6*(September), 1–11. https://doi.org/10.1038/srep30823

Campos-Bermudez, V. A., Bologna, F. P., Andreo, C. S., & Drincovich, M. F. (2010). Functional dissection of Escherichia coli phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation. *FEBS Journal*, *277*(8). https://doi.org/10.1111/j.1742-4658.2010.07617.x

Cao, Z., Casabona, M. G., Kneuper, H., Chalmers, J. D., & Palmer, T. (2016). The type VII secretion system of Staphylococcus aureus secretes a nuclease toxin that targets competitor bacteria. *Nature Microbiology*, *2*(1). https://doi.org/10.1038/nmicrobiol.2016.183

Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., Aarestrup, F. M., & Hasman, H. (2014). In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*, *58*(7). https://doi.org/10.1128/AAC.02412-14

Casjens, S. (2003). Prophages and bacterial genomics: What have we learned so far? In *Molecular Microbiology* (Vol. 49, Issue 2, pp. 277–300). Mol Microbiol. https://doi.org/10.1046/j.1365-2958.2003.03580.x

CGE Server. (n.d.). Retrieved May 8, 2019, from https://cge.cbs.dtu.dk/services/

Chadha, T. (2014). *Bacterial Biofilms: Survival Mechanisms and Antibiotic Resistance*. https://doi.org/10.4172/2155-9597.1000190

Chaibenjawong, P., & Foster, S. J. (2011). Desiccation tolerance in Staphylococcus aureus. *Archives of Microbiology*, *193*(2), 125–135. https://doi.org/10.1007/s00203-010-0653-x

Chaieb, K., Mahdouani, K., & Bakhrouf, A. (2005). Detection of icaA and icaD loci by polymerase chain reaction and biofilm formation by Staphylococcus epidermidis isolated from dialysate and needles in a dialysis unit. *Journal of Hospital Infection*, *61*(3). https://doi.org/10.1016/j.jhin.2005.05.014

Chambers, H. F., & DeLeo, F. R. (2009). Waves of resistance: Staphylococcus aureus in the antibiotic era. In *Nature Reviews Microbiology* (Vol. 7, Issue 9). https://doi.org/10.1038/nrmicro2200

Chamchod, F., & Palittapongarnpim, P. (2019). Effects of the proportion of high-risk patients and control strategies on the prevalence of methicillin-resistant Staphylococcus aureus in an intensive care unit. *BMC Infectious Diseases*, *19*(1). https://doi.org/10.1186/s12879-019-4632-9

Chan, C. X., Beiko, R. G., & Ragan, M. A. (2011). Lateral Transfer of genes and gene fragments in Staphylococcus extends beyond mobile elements. *Journal of Bacteriology*, *193*(15). https://doi.org/10.1128/JB.01524-10

Chatterjee, S. S., & Otto, M. (2013). Improved understanding of factors driving methicillin-resistant Staphylococcus aureus epidemic waves. *Clinical Epidemiology*, *5*(1), 205–217. https://doi.org/10.2147/CLEP.S37071

Chen, H., Lyne, P., Giordanetto, F., Lovell, T. and Li, J. (2005). On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. Journal of Chemical Information and Modeling, 46(1), pp.401-415

Chen, T. R., Chiou, C. S., & Tsen, H. Y. (2004). Use of novel PCR primers specific to the genes of staphylococcal enterotoxin G, H, I for the survey of Staphylococcus aureus strains isolated from food-poisoning cases and food samples in Taiwan. *International Journal of Food Microbiology*, *92*(2), 189–197. https://doi.org/10.1016/j.ijfoodmicro.2003.10.002

Cheng, A. G., McAdow, M., Kim, H. K., Bae, T., Missiakas, D. M., & Schneewind, O. (2010). Contribution of coagulases towards Staphylococcus aureus disease and protective immunity. *PLoS Pathogens*, *6*(8). https://doi.org/10.1371/journal.ppat.1001036

Cho, S. H., Naber, K., Hacker, J., & Ziebuhr, W. (2002). Detection of the icaADBC gene cluster and biofilm formation in Staphylococcus epidermidis isolates from catheter-related urinary tract infections. *International Journal of Antimicrobial Agents*, *19*(6), 570–575. https://doi.org/10.1016/S0924-8579(02)00101-2

Chua, K. Y. L., Howden, B. P., Jiang, J. H., Stinear, T., & Peleg, A. Y. (2014a). Population genetics and the evolution of virulence in Staphylococcus aureus. *Infection, Genetics and Evolution*, *21*, 554–562. https://doi.org/10.1016/j.meegid.2013.04.026

Clokie, M. R. J., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31–45. https://doi.org/10.4161/bact.1.1.14942

CLSI. (2014). *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fourth Informational Supplement* (9th ed., Vol. 34).

Colavecchio, A., Cadieux, B., Lo, A., & Goodridge, L. D. (2017). Bacteriophages Contribute to the Spread of Antibiotic Resistance Genes among Foodborne Pathogens of the Enterobacteriaceae Family – A Review. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.01108

Collee, J., Fraser, A., & Marmion, B. P. (1996). *Mackie and McCartney Practical Medical Microbiology* (14th ed.). Elsevier Health Sciences.

Coombs, G. W., Baines, S. L., Howden, B. P., Swenson, K. M., & O'Brien, F. G. (2020). Diversity of bacteriophages encoding Panton-Valentine leukocidin in temporally and geographically related Staphylococcus aureus. *PLOS ONE*, *15*(2). https://doi.org/10.1371/journal.pone.0228676

Consortium, T. U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., … Teodoro, D. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/NAR/GKAA1100

Cramton, S. E., Gerke, C., Schnell, N. F., Nichols, W. W., & Götz, F. (1999). The intercellular adhesion (ica) locus is present in Staphylococcus aureus and is required for biofilm formation. *Infection and Immunity*, *67*(10). https://doi.org/10.1128/iai.67.10.5427-5433.1999

Cumby, N., Davidson, A. R., & Maxwell, K. L. (2012). The moron comes of age. *Bacteriophage*, *2*(4), e23146. https://doi.org/10.4161/bact.23146

Daina, A., Michielin, O. and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness, and medicinal chemistry friendliness of small molecules. Scientific Reports, 7(1)

Darboe, S., Dobreniecki, S., Jarju, S., Jallow, M., Mohammed, N. I., Wathuo, M., Ceesay, B., Tweed, S., Roy, R. B., Okomo, U., Kwambana-Adams, B., Antonio, M., Bradbury, R. S., de Silva, T. I., Forrest, K., Roca, A., Lawal, B. J., Nwakanma, D., & Secka, O. (2019). Prevalence of Panton-Valentine Leukocidin (PVL) and Antimicrobial Resistance in Community-Acquired Clinical Staphylococcus aureus in an Urban Gambian Hospital: A

11-year period retrospective pilot study. *Frontiers in Cellular and Infection Microbiology*, *9*(MAY). https://doi.org/10.3389/fcimb.2019.00170

Darling, A. E., Mau, B., & Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0011147

de Carvalho, S. P., de Almeida, J. B., de Freitas, L. M., Guimarães, A. M. S., do Nascimento, N. C., dos Santos, A. P., Campos, G. B., Messick, J. B., Timenetsky, J., & Marques, L. M. (2019). Genomic profile of Brazilian methicillin-resistant staphylococcus aureus resembles clones dispersed worldwide. *Journal of Medical Microbiology*, *68*(5), 693–702. https://doi.org/10.1099/jmm.0.000956

De Paepe, M., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S., & Petit, M.-A. (2014). Temperate Phages Acquire DNA from Defective Prophages by Relaxed Homologous Recombination: The Role of Rad52-Like Recombinases. *PLoS Genetics*, *10*(3), e1004181. https://doi.org/10.1371/journal.pgen.1004181

Deghorain, M., & van Melderen, L. (2012). The Staphylococci Phages Family: An Overview. *Viruses*, *4*(12), 3316–3335. https://doi.org/10.3390/v4123316

DeLeo, F. R., Diep, B. A., & Otto, M. (2009). Host Defense and Pathogenesis in Staphylococcus aureus Infections. In *Infectious Disease Clinics of North America* (Vol. 23, Issue 1). https://doi.org/10.1016/j.idc.2008.10.003

Denis, O. (2017). Route of transmission of Staphylococcus aureus. *The Lancet Infectious Diseases*, *17*(2), 124–125. https://doi.org/10.1016/S1473-3099(16)30512-6

Denis, O., Nashev, D., Blanc, D. S., Pieridou-Bagatzouni, D., Jakubu, V., Zemlickova, H., Westh, H., Larsen, A. R., Skov, R., Laurent, F., Layer, F., Witte, W., Spiliopoulou, I., Salmenlinna, S., Lindholm, L., Vuopio-Varkila, J., Toth, A., Ungvari, E., Brennan, G., … Edwards, G. (2014). The dynamic changes of dominant clones of Staphylococcus aureus causing bloodstream infections in the European region: Results of a second structured survey. *Eurosurveillance*, *19*(49). https://doi.org/10.2807/1560-7917.ES2014.19.49.20987

Deurenberg, R. H., Vink, C., Kalenic, S., Friedrich, A. W., Bruggeman, C. A., & Stobberingh, E. E. (2007). The molecular evolution of methicillin-resistant Staphylococcus aureus. In *Clinical Microbiology and Infection* (Vol. 13, Issue 3, pp. 222–235). Blackwell Publishing Ltd. https://doi.org/10.1111/j.1469-0691.2006.01573.x

Deutscher, J., Herro, R., Bourand, A., Mijakovic, I., & Poncet, S. (2005). P-Ser-HPr - A link between carbon metabolism and the virulence of some pathogenic bacteria. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1754*(1–2), 118–125. https://doi.org/10.1016/j.bbapap.2005.07.029

Dinges, M. M., Orwin, P. M., & Schlievert, P. M. (2000). Exotoxins of Staphylococcus aureus. *Clinical Microbiology Reviews*, *13*(1). https://doi.org/10.1128/CMR.13.1.16-34.2000

Discriminatory power Calculator. (n.d.). Retrieved February 8, 2019, from http://insilico.ehu.es/mini_tools/discriminatory_power/

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-4

Du, J., Chen, C., Ding, B., Tu, J., Qin, Z., Parsons, C., Salgado, C., Cai, Q., Song, Y., Bao, Q., Zhang, L., Pan, J., Wang, L., & Yu, F. (2011). Molecular Characterization and Antimicrobial Susceptibility of Nasal Staphylococcus aureus Isolates from a Chinese Medical College Campus. *PLoS ONE*, *6*(11), e27328. https://doi.org/10.1371/journal.pone.0027328

Duffield, M., Cooper, I., McAlister, E., Bayliss, M., Ford, D., & Oyston, P. (2010). Predicting conserved essential genes in bacteria: in silico identification of putative drug targets. *Molecular BioSystems*, *6*(12). https://doi.org/10.1039/c0mb00001a

DuMont, A. L., Yoong, P., Surewaard, B. G. J., Benson, M. A., Nijland, R., Strijp, J. A. G. van, & Torres, V. J. (2013). Staphylococcus aureus elaborates leukocidin AB to mediate escape from within human neutrophils. *Infection and Immunity*, *81*(5). https://doi.org/10.1128/IAI.00095-13

Eckhart, L., Fischer, H., Barken, K. B., Tolker-Nielsen, T., & Tschachler, E. (2007). DNase1L2 suppresses biofilm formation by Pseudomonas aeruginosa and Staphylococcus aureus. *British Journal of Dermatology*, *156*(6), 1342–1345. https://doi.org/10.1111/j.1365-2133.2007.07886.x

el Farran, C. A., Sekar, A., Balakrishnan, A., Shanmugam, S., Arumugam, P., & Gopalswamy, J. (2013). Prevalence of biofilm-producing Staphylococcus epidermidis in the healthy skin of individuals in Tamil Nadu, India. *Indian Journal of Medical Microbiology*, *31*(1). https://doi.org/10.4103/0255-0857.108712

Enright, M. C., & Spratt, B. G. (1999). Multilocus sequence typing. In *Trends in Microbiology* (Vol. 7, Issue 12). https://doi.org/10.1016/S0966-842X(99)01609-1

Fairbrother, R. W. (1940). Coagulase production as a criterion for the classification of the staphylococci. *The Journal of Pathology and Bacteriology*, *50*(1). https://doi.org/10.1002/path.1700500112

Farris, J. S. (1972). Estimating Phylogenetic Trees from Distance Matrices. *The American Naturalist*, *106*(951). https://doi.org/10.1086/282802

Feng, Y., Chen, C.-J., Su, L.-H., Hu, S., Yu, J., & Chiu, C.-H. (2008a). Evolution and pathogenesis of Staphylococcus aureus : lessons learned from genotyping and comparative genomics. *FEMS Microbiology Reviews*, *32*(1), 23–37. https://doi.org/10.1111/j.1574-6976.2007.00086.x

Fitzgerald, J. R. (2012). Livestock-associated Staphylococcus aureus: Origin, evolution and public health threat. In *Trends in Microbiology* (Vol. 20, Issue 4, pp. 192–198). Trends Microbiol. https://doi.org/10.1016/j.tim.2012.01.006

Fitzgerald, J. R., & Holden, M. T. G. (2016). Genomics of Natural Populations of Staphylococcus aureus. *Annual Review of Microbiology*, *70*. https://doi.org/10.1146/annurev-micro-102215-095547

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R., & Musser, J. M. (2001). Evolutionary genomics of Staphylococcus aureus: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 8821–8826. https://doi.org/10.1073/pnas.161098098

Fitzpatrick, F., Humphreys, H., & O'Gara, J. P. (2005). Evidence for icaADBC-Independent Biofilm Development Mechanism in Methicillin-Resistant Staphylococcus aureus Clinical Isolates. *Journal of Clinical Microbiology*, *43*(4), 1973–1976. https://doi.org/10.1128/JCM.43.4.1973-1976.2005

Fluit, A. C. (2012). Livestock-associated Staphylococcus aureus. *Clinical Microbiology and Infection*, *18*(8), 735–744. https://doi.org/10.1111/j.1469-0691.2012.03846.x

Fortier, L. C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

Foster. (1996). Staphylococcus. In Baron S (Ed.), *Medical Microbiology* (4th ed.). Galveston (TX): University of Texas Medical Branch at Galveston. https://www.ncbi.nlm.nih.gov/books/NBK8448/

Foster, T. J. (2004). The Staphylococcus aureus "superbug." *Journal of Clinical Investigation*, *114*(12), 1693–1696. https://doi.org/10.1172/jci23825

Foster, T. J., Geoghegan, J. A., Ganesh, V. K., & Höök, M. (2014). Adhesion, invasion and evasion: The many functions of the surface proteins of Staphylococcus aureus. In *Nature Reviews Microbiology* (Vol. 12, Issue 1). https://doi.org/10.1038/nrmicro3161

Freney, J., Brun, Y., Bes, M., Meugnier, H., Grimont, F., Grimont, P. A. D., Nervi, C., & Fleurette, J. (1988). Staphylococcus lugdunensis sp. nov. and Staphylococcus schleiferi sp. nov., two species from human clinical specimens. *International Journal of Systematic Bacteriology*, *38*(2). https://doi.org/10.1099/00207713-38-2-168

Galata, V., Fehlmann, T., Backes, C., & Keller, A. (2019). PLSDB: A resource of complete bacterial plasmids. *Nucleic Acids Research*, *47*(D1), D195–D202. https://doi.org/10.1093/nar/gky1050

Galdbart, J., Allignet, J., Tung, H., Rydèn, C., & El Solh, N. (2000). Screening for Staphylococcus epidermidis Markers Discriminating between Skin-Flora Strains and Those Responsible for Infections of Joint Prostheses. *The Journal of Infectious Diseases*, *182*(1), 351–355. https://doi.org/10.1086/315660

Gascuel, O. (1997). BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, *14*(7), 685–695. https://doi.org/10.1093/oxfordjournals.molbev.a025808

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563

Genevaux, P., Bauda, P., DuBow, M. S., & Oudega, B. (1999). Identification of Tn 10 insertions in the rlaG, rfaP, and gaLU genes involved in lipopolysaccharide core

biosynthesis that affect Escherichia coli adhesion. *Archives of Microbiology*, *172*(1), 1–8. https://doi.org/10.1007/s002030050732

Ghasemian, A., Peerayeh, S. N., Bakhshi, B., & Mirzaee, M. (2015). The microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) genes among clinical isolates of staphylococcus aureus from hospitalized children. *Iranian Journal of Pathology*, *10*(4). https://doi.org/10.7508/ijp.2015.04.002

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J., Khouri, H., Utterback, T., … Fraser, C. M. (2005). Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant Staphylococcus aureus Strain and a Biofilm-Producing Methicillin-Resistant Staphylococcus epidermidis Strain. *Journal of Bacteriology*, *187*(7), 2426–2438. https://doi.org/10.1128/JB.187.7.2426-2438.2005

Goerke, C., Koller, J., & Wolz, C. (2006). Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *50*(1), 171–177. https://doi.org/10.1128/AAC.50.1.171-177.2006

Goh, S. H., Byrne, S. K., Zhang, J. L., & Chow, A. W. (1992). Molecular typing of Staphylococcus aureus on the basis of coagulase gene polymorphisms. *Journal of Clinical Microbiology*, *30*(7), 1642–1645.

Goncearenco, A., Li, M., Simonetti, F. L., Shoemaker, B. A., & Panchenko, A. R. (2017). Exploring protein-protein interactions as drug targets for anti-cancer therapy with in silico workflows. In *Methods in Molecular Biology* (Vol. 1647, pp. 221–236). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7201-2_15

Gordon, E., Flouret, B., Chantalat, L., van Heijenoort, J., Mengin-Lecreulx, D., & Dideberg, O. (2001). Crystal Structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-Diaminopimelate Ligase from Escherichia Coli. *Journal of Biological Chemistry*, *276*(14), 10999–11006. https://doi.org/10.1074/jbc.M009835200

Gordon, R. J., & Lowy, F. D. (2008). Pathogenesis of Methicillin-Resistant Staphylococcus aureus Infection. *Clinical Infectious Diseases*, *46*(S5), S350–S359. https://doi.org/10.1086/533591

Götz, F., Bannerman, T., & Schleifer, K.-H. (2006). The Genera Staphylococcus and Macrococcus. In *The Prokaryotes* (pp. 5–75). Springer US. https://doi.org/10.1007/0-387-30744-3_1

Goudarzi, M., Seyedjavadi, S. S., Nasiri, M. J., Goudarzi, H., Sajadi Nia, R., & Dabiri, H. (2017). Molecular characteristics of methicillin-resistant Staphylococcus aureus (MRSA) strains isolated from patients with bacteremia based on MLST, SCCmec, spa, and agr locus types analysis. *Microbial Pathogenesis*, *104*, 328–335. https://doi.org/10.1016/j.micpath.2017.01.055

Goyal, M., & Singh, N. (2018). In silico identification of novel drug targets in acinetobacter baumannii by subtractive genomic approach. *Article in Asian Journal of Pharmaceutical and Clinical Research*, *11*. https://doi.org/10.22159/ajpcr.2018.v11i3.22105

Grumann, D., Nübel, U., & Bröker, B. M. (2014). Staphylococcus aureus toxins - Their functions and genetics. *Infection, Genetics and Evolution*, *21*. https://doi.org/10.1016/j.meegid.2013.03.013

Grundmann, H., Aanensen, D. M., van den Wijngaard, C. C., Spratt, B. G., Harmsen, D., Friedrich, A. W., Sabat, A. J., Muilwijk, J., Monen, J., Tami, A., Donker, T., Mittermayer, H., Krziwanek, K., Stumvoll, S., Koller, W., Denis, O., Struelens, M., Nashev, D., Budimir, A., … Vogel, U. (2010). Geographic distribution of Staphylococcus aureus causing invasive infections in Europe: A molecular-epidemiological analysis. *PLoS Medicine*, *7*(1). https://doi.org/10.1371/journal.pmed.1000215

Gupta, E., Mishra, R. K., & Niraj, R. R. K. (2020). Identification of potential vaccine candidates against SARS-CoV-2, a step forward to fight novel coronavirus 2019-nCoV: A reverse vaccinology approach. In *bioRxiv* (p. 2020.04.13.039198). bioRxiv. https://doi.org/10.1101/2020.04.13.039198

Guttman, B., Raya, R., Kutter, E., Kutter, E., & Sulakvelidze, A. (2005). Basic phage biology. In E. Kutter & A. Sulakvelidze (Eds.), *Bacteriophages: biology and applications* (pp. 29–66). CRC Press. Gutiérrez D, Hidalgo-Cantabrana C, Rodríguez A, García P, Ruas-Madiedo P (2016) Monitoring in Real Time the Formation and Removal of Biofilms from Clinical Related Pathogens Using an Impedance-Based Technology. PLoS ONE 11(10): e0163966. https://doi.org/10.1371/journal.pone.0163966

Hacker, J., & Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Reports*, *2*(5). https://doi.org/10.1093/embo-reports/kve097

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). Past: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*(1).

Harmsen, D., Claus, H., Witte, W., Rothgänger, J., Claus, H., Turnwald, D., & Vogel, U. (2003). Typing of Methicillin-Resistant Staphylococcus aureus in a University Hospital Setting by Using Novel Software for spa Repeat Determination and Database Management. *Journal of Clinical Microbiology*, *41*(12). https://doi.org/10.1128/JCM.41.12.5442-5448.2003

Hayek, N. (2013). Lateral transfer and GC content of bacterial resistance genes. In *Frontiers in Microbiology* (Vol. 4, Issue MAR). https://doi.org/10.3389/fmicb.2013.00041

He, Y., Xiang, Z., & Mobley, H. L. T. (2010). Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Journal of Biomedicine and Biotechnology*, *2010*. https://doi.org/10.1155/2010/297505

Hiramatsu, K., Cui, L., Kuroda, M., & Ito, T. (2001). The emergence and evolution of methicillin-resistant Staphylococcus aureus. In *Trends in Microbiology* (Vol. 9, Issue 10). https://doi.org/10.1016/S0966-842X(01)02175-8

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T., & Tenover, F. C. (1997). Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility

[1]. In *Journal of Antimicrobial Chemotherapy* (Vol. 40, Issue 1). https://doi.org/10.1093/jac/40.1.135

Hiramatsu, K., Katayama, Y., Matsuo, M., Sasaki, T., Morimoto, Y., Sekiguchi, A., & Baba, T. (2014). Multi-drug-resistant Staphylococcus aureus and future chemotherapy. In *Journal of Infection and Chemotherapy* (Vol. 20, Issue 10). https://doi.org/10.1016/j.jiac.2014.08.001

Hoang, T. M., Zhou, C., Lindgren, J. K., Galac, M. R., Corey, B., Endres, J. E., Olson, M. E., & Fey, P. D. (2019). Transcriptional regulation of icaADBC by both IcaR and TcaR in staphylococcus epidermidis. *Journal of Bacteriology*, *201*(6). https://doi.org/10.1128/JB.00524-18

Holden, M. T. G., Hsu, L. Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Žemličková, H., Coombs, G., Kearns, A. M., Hill, R. L. R., Edgeworth, J., Gould, I., Gant, V., … Nübel, U. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Research*, *23*(4). https://doi.org/10.1101/gr.147710.112

Holmes, A., Ganner, M., McGuane, S., Pitt, T. L., Cookson, B. D., & Kearns, A. M. (2005). Staphylococcus aureus Isolates Carrying Panton-Valentine Leucocidin Genes in England and Wales: Frequency, Characterization, and Association with Clinical Disease. *Journal of Clinical Microbiology*, *43*(5), 2384–2390. https://doi.org/10.1128/JCM.43.5.2384-2390.2005

Hookey, J. V, Richardson, J. F., & Cookson, B. D. (1998). Molecular typing of Staphylococcus aureus based on PCR restriction fragment length polymorphism and DNA sequence analysis of the coagulase gene. *Journal of Clinical Microbiology*, *36*(4), 1083–1089.

Hu, D. L., & Nakane, A. (2014). Mechanisms of staphylococcal enterotoxin-induced emesis. In *European Journal of Pharmacology* (Vol. 722, Issue 1). https://doi.org/10.1016/j.ejphar.2013.08.050

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1085

Hunter, P. R., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. *Journal of Clinical Microbiology*, *26*(11). https://doi.org/10.1128/jcm.26.11.2465-2466.1988

Huseby, M. J., Kruse, A. C., Digre, J., Kohler, P. L., Vocke, J. A., Mann, E. E., Bayles, K. W., Bohach, G. A., Schlievert, P. M., Ohlendorf, D. H., & Earhart, C. A. (2010). Beta toxin catalyzes formation of nucleoprotein matrix in staphylococcal biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32). https://doi.org/10.1073/pnas.0911032107

Huseby, M., Shi, K., Kent Brown, C., Digre, J., Mengistu, F., Keun, S. S., Bohach, G. A., Schlievert, P. M., Ohlendorf, D. H., & Earhart, C. A. (2007). Structure and biological activities of beta toxin from Staphylococcus aureus. *Journal of Bacteriology*, *189*(23). https://doi.org/10.1128/JB.00741-07

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. In *Molecular Biology and Evolution* (Vol. 23, Issue 2, pp. 254–267). Mol Biol Evol. https://doi.org/10.1093/molbev/msj030

Hussain, M., von Eiff, C., Sinha, B., Joost, I., Herrmann, M., Peters, G., & Becker, K. (2008). eap Gene as Novel Target for Specific Identification of Staphylococcus aureus. *Journal of Clinical Microbiology*, *46*(2), 470–476. https://doi.org/10.1128/JCM.01425-07

Ito, T., Hiramatsu, K., Oliveira, D. C., de Lencastre, H., Zhang, K., Westh, H., O'Brien, F., Giffard, P. M., Coleman, D., Tenover, F. C., Boyle-Vavra, S., Skov, R. L., Enright, M. C., Kreiswirth, B., Kwan, S. K., Grundmann, H., Laurent, F., Sollid, J. E., Kearns, A. M., … Soderquist, B. (2009). Classification of staphylococcal cassette chromosome mec (SCCmec): Guidelines for reporting novel SCCmec elements. In *Antimicrobial Agents and Chemotherapy* (Vol. 53, Issue 12, pp. 4961–4967). Antimicrob Agents Chemother. https://doi.org/10.1128/AAC.00579-09

Ito, T., Kuwahara-Arai, K., Katayama, Y., Uehara, Y., Han, X., Kondo, Y., & Hiramatsu, K. (2014). Staphylococcal cassette chromosome mec (SCCmec) analysis of MRSA. *Methods in Molecular Biology*, *1085*, 131–148. https://doi.org/10.1007/978-1-62703-664-1_8

Jahid, I., Lee, N., Kim, A., & Ha, S. (2013). Influence of Glucose Concentrations on Biofilm Formation, Motility, Exoprotease Production, and Quorum Sensing in Aeromonas hydrophila. *Journal Of Food Protection*, *76*(2), 239-247. https://doi.org/10.4315/0362-028x.jfp-12-321

Janwithayanuchit, I., Ngam-Ululert, S., Paungmoung, P., & Rangsipanuratn, W. (2006). Epidemiologic Study of methicillin-resistant Staphylococcus aureus by Coagulase Gene Polymorphism. *ScienceAsia*, *32*, 127–132. https://doi.org/10.2306/scienceasia1513-1874.2006.32.127

Jarraud, S., Cozon, G., Vandenesch, F., Bes, M., Etienne, J., & Lina, G. (1999). Involvement of enterotoxins G and I in staphylococcal toxic shock syndrome and staphylococcal scarlet fever. *Journal of Clinical Microbiology*, *37*(8). https://doi.org/10.1128/jcm.37.8.2446-2449.1999

Jarraud, S., Peyrat, M. A., Lim, A., Tristan, A., Bes, M., Mougel, C., Etienne, J., Vandenesch, F., Bonneville, M., & Lina, G. (2001). egc , A Highly Prevalent Operon of Enterotoxin Gene, Forms a Putative Nursery of Superantigens in Staphylococcus aureus . *The Journal of Immunology*, *166*(1). https://doi.org/10.4049/jimmunol.166.1.669

Jiang, X., Yang, Y., Zhou, J., Liu, H., Liao, X., Luo, J., Li, X., & Fang, W. (2019). Peptidyl isomerase PrsA is surface-associated on Streptococcus suis and offers cross-protection against serotype 9 strain. *FEMS Microbiology Letters*, *366*(2), 2. https://doi.org/10.1093/FEMSLE/FNZ002

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., & Aarestrup, F. M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *Journal of Clinical Microbiology*, *52*(5), 1501–1510. https://doi.org/10.1128/JCM.03617-13

Juhas, M., Eberl, L., & Church, G. M. (2012). Essential genes as antimicrobial targets and cornerstones of synthetic biology. In *Trends in Biotechnology* (Vol. 30, Issue 11, pp. 601–607). Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2012.08.002

Kala, S., Cumby, N., Sadowski, P. D., Hyder, B. Z., Kanelis, V., Davidson, A. R., & Maxwell, K. L. (2014). HNH proteins are a widespread component of phage DNA packaging machines. *Proceedings of the National Academy of Sciences*, *111*(16), 6022–6027. https://doi.org/10.1073/pnas.1320952111

Kaneko, J., Kimura, T., Narita, S., Tomita, T., & Yoshiyuki Kamio. (1998). Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage φPVL carrying Panton–Valentine leukocidin genes. *Gene*, *215*(1), 57–67. https://doi.org/10.1016/S0378-1119(98)00278-9

Kareem, S. M., Aljubori, S. S., & Ali, M. R. (2020). Novel determination of spa gene diversity and its molecular typing among Staphylococcus aureus Iraqi isolates obtained from different clinical samples. *New Microbes and New Infections*, *34*. https://doi.org/10.1016/j.nmni.2020.100653

Katayama, Y., Baba, T., Sekine, M., Fukuda, M., & Hiramatsu, K. (2013). Beta-hemolysin promotes skin colonization by Staphylococcus aureus. *Journal of Bacteriology*, *195*(6). https://doi.org/10.1128/JB.01786-12

Katayama, Y., Ito, T., & Hiramatsu, K. (2000). A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *44*(6). https://doi.org/10.1128/AAC.44.6.1549-1555.2000

Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4). https://doi.org/10.1093/bib/bbx108

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., Lemvigh, C. K., Aarestrup, F. M., Lund, O., & Larsen, A. R. (2018). SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. *MSphere*, *3*(1). https://doi.org/10.1128/msphere.00612-17

Keating, T. A., Newman, J. v., Olivier, N. B., Otterson, L. G., Andrews, B., Boriack-Sjodin, P. A., Breen, J. N., Doig, P., Dumas, J., Gangl, E., Green, O. M., Guler, S. Y., Hentemann, M. F., Joseph-McCarthy, D., Kawatkar, S., Kutschke, A., Loch, J. T., McKenzie, A. R., Pradeepan, S., … Martínez-Botella, G. (2012). In Vivo Validation of Thymidylate Kinase (TMK) with a Rationally Designed, Selective Antibacterial Compound. *ACS Chemical Biology*, *7*(11). https://doi.org/10.1021/cb300316n

Khalida, S., Nadia, R., & Mohammad, S. (2012). Subtractive genome analysis for in silico identification and characterization of novel drug targets in c. trachomatis strain d/uw-3/cx. *International Journal of Current Research*, *4*(7), 017–021.

Khoshkharam-Roodmajani, H., Sarvari, J., Bazargani, A., Kandekar-Ghahraman, M. R., Nazari-Alam, A., & Motamedifar, M. (2014). Molecular typing of methicillin-resistant and methicillin-susceptible Staphylococcus aureus isolates from Shiraz teaching hospitals by PCR-RFLP of coagulase gene. *Iranian Journal of Microbiology*, *6*(4), 246–252. http://ijm.tums.ac.ir

Kiem, S., Oh, W. S., Peck, K. R., Lee, N. Y., Lee, J. Y., Song, J. H., Hwang, E. S., Kim, E. C., Cha, C. Y., & Choe, K. W. (2004). Phase variation of biofilm formation in Staphylococcus aureus by IS256 insertion and its impact on the capacity adhering to polyurethane surface. *Journal of Korean Medical Science*, *19*(6), 779–782. https://doi.org/10.3346/jkms.2004.19.6.779

Kim, J., Gambhir, V., Alatery, A., & Basta, S. (2010). Delivery of exogenous antigens to induce cytotoxic CD8+ T lymphocyte responses. *Journal of Biomedicine and Biotechnology*, *2010*. https://doi.org/10.1155/2010/218752

Kim, Y., Li, H., Binkowski, T. A., Holzle, D., & Joachimiak, A. (2009). Crystal structure of fatty acid/phospholipid synthesis protein PlsX from Enterococcus faecalis. *Journal of Structural and Functional Genomics*, *10*(2). https://doi.org/10.1007/s10969-008-9052-9

Kleinert, F., Kallies, R., Zweynert, A., & Bierbaum, G. (2016). Draft genome sequences of three northern German epidemic Staphylococcus aureus (ST247) strains containing multiple copies of IS256. *Genome Announcements*, *4*(5). https://doi.org/10.1128/genomeA.00936-16

Kloos, W. E., & Bannerman, T. L. (1994). Update on clinical significance of coagulase-negative staphylococci. In *Clinical Microbiology Reviews* (Vol. 7, Issue 1). https://doi.org/10.1128/CMR.7.1.117

Kluytmans, J., van Belkum, A., & Verbrugh, H. (1997). Nasal carriage of Staphylococcus aureus: Epidemiology, underlying mechanisms, and associated risks. In *Clinical Microbiology Reviews* (Vol. 10, Issue 3, pp. 505–520). American Society for Microbiology. https://doi.org/10.1128/cmr.10.3.505-520.1997

Köck, R., Becker, K., Cookson, B., van Gemert-Pijnen, J. E., Harbarth, S., Kluytmans, J., Mielke, M., Peters, G., Skov, R. L., Struelens, M. J., Tacconelli, E., Torné, A. N., Witte, W., & Friedrich, A. W. (2010). Methicillin-resistant Staphylococcus aureus (MRSA): Burden of disease and control challenges in Europe. In *Eurosurveillance* (Vol. 15, Issue 41, pp. 1–9). European Centre for Disease Prevention and Control (ECDC). https://doi.org/10.2807/ese.15.41.19688-en

Koreen, L., Ramaswamy, S. V, Graviss, E. A., Naidich, S., Musser, J. M., & Kreiswirth, B. N. (2004). spa typing method for discriminating among Staphylococcus aureus isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *Journal of Clinical Microbiology*, *42*(2), 792–799. https://doi.org/10.1128/jcm.42.2.792-799.2004

Köser, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T. G., Dougan, G., Bentley, S. D., Parkhill, J., & Peacock, S. J. (2012). Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. *PLoS Pathogens*, *8*(8), e1002824. https://doi.org/10.1371/journal.ppat.1002824

Kumar Jaiswal, A., Tiwari, S., Jamal, S., Barh, D., Azevedo, V., & Soares, S. (2017). An In Silico Identification of Common Putative Vaccine Candidates against Treponema pallidum: A Reverse Vaccinology and Subtractive Genomics Based Approach. *International Journal of Molecular Sciences*, *18*(2), 402. https://doi.org/10.3390/ijms18020402

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Kobayashi, N., Cui, L., Oguchi, A., Aoki, K. I., Nagai, Y., Lian, J. Q., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., … Hiramatsu, K. (2001). Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet*, *357*(9264), 1225–1240. https://doi.org/10.1016/S0140-6736(00)04403-2

Kwan, T., Liu, J., DuBow, M., Gros, P., & Pelletier, J. (2005). The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(14), 5174–5179. https://doi.org/10.1073/pnas.0501140102

Kwong, J. C., Mccallum, N., Sintchenko, V., & Howden, B. P. (2015). Whole genome sequencing in clinical and public health microbiology. *Pathology*, *47*(3), 199–210. https://doi.org/10.1097/PAT.0000000000000235

LaBreck, P. T., Rice, G. K., Paskey, A. C., Elassal, E. M., Cer, R. Z., Law, N. N., Schlett, C. D., Bennett, J. W., Millar, E. v., Ellis, M. W., Hamilton, T., Bishop-Lilly, K. A., & Scott Merrell, D. (2018). Conjugative transfer of a novel staphylococcal plasmid encoding the biocide resistance gene, QacA. *Frontiers in Microbiology*, *9*(NOV). https://doi.org/10.3389/fmicb.2018.02664

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. In *Nature Reviews Microbiology* (Vol. 8, Issue 5, pp. 317–327). Nat Rev Microbiol. https://doi.org/10.1038/nrmicro2315

Ladhani, S. (2001). Recent developments in staphylococcal scalded skin syndrome. In *Clinical Microbiology and Infection* (Vol. 7, Issue 6). https://doi.org/10.1046/j.1198-743X.2001.00258.x

Ladhani, S. (2003). Understanding the mechanism of action of the exfoliative toxins of Staphylococcus aureus. In *FEMS Immunology and Medical Microbiology* (Vol. 39, Issue 2). https://doi.org/10.1016/S0928-8244(03)00225-6

Laing, C., Buchanan, C., Taboada, E. N., Zhang, Y., Kropinski, A., Villegas, A., Thomas, J. E., & Gannon, V. P. J. (2010). Pan-genome sequence analysis using Panseq: An online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-461

Larsen, M. v., Cosentino, S., Lukjancenko, O., Saputra, D., Rasmussen, S., Hasman, H., Sicheritz-Pontén, T., Aarestrup, F. M., Ussery, D. W., & Lund, O. (2014). Benchmarking of methods for genomic taxonomy. *Journal of Clinical Microbiology*, *52*(5), 1529–1539. https://doi.org/10.1128/JCM.02981-13

Larsen, M. v., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D. W., Aarestrup, F. M., & Lund, O. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology*, *50*(4), 1355–1361. https://doi.org/10.1128/JCM.06094-11

Lazarevic, V., Beaume, M., Corvaglia, A., Hernandez, D., Schrenzel, J., & François, P. (2011). Epidemiology and virulence insights from MRSA and MSSA genome analysis. In *Future Microbiology* (Vol. 6, Issue 5). https://doi.org/10.2217/fmb.11.38

Lee, C. Y., & Iandolo, J. J. (1986). Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. *Journal of Bacteriology*, *166*(2), 385–391. https://doi.org/10.1128/JB.166.2.385-391.1986

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*(10). https://doi.org/10.1093/molbev/msv150

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Research*, *47*(W1), W260–W265. https://doi.org/10.1093/nar/gkz303

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1). https://doi.org/10.1093/nar/gkz239

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Licitra, G. (2013). Etymologia: Staphylococcus . *Emerging Infectious Diseases*, *19*(9), 1553. https://doi.org/10.3201/eid1909.et1909

Lim, S., Lee, D. H., Kwak, W., Shin, H., Ku, H. J., Lee, J. E., Lee, G. E., Kim, H., Choi, S. H., Ryu, S., & Lee, J. H. (2014). Comparative genomic analysis of Staphylococcus aureus FORC_001 and S. aureus MRSA252 reveals the Characteristics of antibiotic resistance and virulence factors for human infection. *Journal of Microbiology and Biotechnology*, *25*(1), 98–108. https://doi.org/10.4014/jmb.1410.10005

Lin, C. F., Chen, C. L., Huang, W. C., Cheng, Y. L., Hsieh, C. Y., Wang, C. Y., & Hong, M. Y. (2010). Different types of cell death induced by Enterotoxins. In *Toxins* (Vol. 2, Issue 8). https://doi.org/10.3390/toxins2082158

Lina, G., Piémont, Y., Godail-Gamot, F., Bes, M., Peter, M. O., Gauduchon, V., Vandenesch, F., & Etienne, J. (1999). Involvement of Panton-Valentine leukocidin-producing Staphylococcus aureus in primary skin infections and pneumonia. *Clinical Infectious Diseases*, *29*(5). https://doi.org/10.1086/313461

Lindsay, J. A. (2010). Genomic variation and evolution of Staphylococcus aureus. In *International Journal of Medical Microbiology* (Vol. 300, Issues 2–3). https://doi.org/10.1016/j.ijmm.2009.08.013

Lindsay, J. A., & Holden, M. T. G. (2006). Understanding the rise of the superbug: Investigation of the evolution and genomic variation of Staphylococcus aureus. In *Functional and Integrative Genomics* (Vol. 6, Issue 3, pp. 186–201). Funct Integr Genomics. https://doi.org/10.1007/s10142-005-0019-7

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D., & Hinds, J. (2006). Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. *Journal of Bacteriology*, *188*(2). https://doi.org/10.1128/JB.188.2.669-676.2006

Liu, D. (2014). Enterotoxin-Producing Staphylococcus aureus. In *Molecular Medical Microbiology: Second Edition* (Vols. 2–3, pp. 979–995). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-397169-2.00055-X

Lowy, F. D. (2003). Antimicrobial resistance: The example of Staphylococcus aureus. In *Journal of Clinical Investigation* (Vol. 111, Issue 9, pp. 1265–1273). The American Society for Clinical Investigation. https://doi.org/10.1172/JCI18535

Lu, Y. J., Zhang, Y. M., Grimes, K. D., Qi, J., Lee, R. E., & Rock, C. O. (2006). Acyl-Phosphates Initiate Membrane Phospholipid Synthesis in Gram-Positive Pathogens. *Molecular Cell*, *23*(5). https://doi.org/10.1016/j.molcel.2006.06.030

Lucchini, S., Desiere, F., & Brüssow, H. (1999). Comparative Genomics of Streptococcus thermophilus Phage Species Supports a Modular Evolution Theory. *Journal of Virology*, *73*(10), 8647–8656. https://doi.org/10.1128/JVI.73.10.8647-8656.1999

Mack, D., Becker, P., Chatterjee, I., Dobinsky, S., Knobloch, J. K., Peters, G., Rohde, H., & Herrmann, M. (2004). Mechanisms of biofilm formation in Staphylococcus epidermidis and Staphylococcus aureus: functional molecules, regulatory circuits, and adaptive responses. International Journal of Medical Microbiology: IJMM, 294(2-3), 203–212. https://doi.org/10.1016/j.ijmm.2004.06.015

Mack, D., Fischer, W., Krokotsch, A., Leopold, K., Hartmann, R., Egge, H., & Laufs, R. (1996). The intercellular adhesin involved in biofilm accumulation of Staphylococcus epidermidis is a linear β-1,6-linked glucosaminoglycan: Purification and structural analysis. *Journal of Bacteriology*, *178*(1). https://doi.org/10.1128/jb.178.1.175-183.1996

Maddux, M. S. (1991). Effects of beta-lactamase-mediated antimicrobial resistance: the role of beta-lactamase inhibitors. *Pharmacotherapy*, *11*(2(2)), 40S-50S.

Mahmoudi, H., Arabestani, M. R., Mousavi, S. F., & Alikhani, M. Y. (2017). Molecular analysis of the coagulase gene in clinical and nasal carrier isolates of methicillin-resistant Staphylococcus aureus by restriction fragment length polymorphism. *Journal of Global Antimicrobial Resistance*, *8*, 41–45. https://doi.org/10.1016/j.jgar.2016.10.007

Maiti, P., Chatterjee, S., Dey, R., Kundu, A., & Dey, R. (2014). Biofilms on indwelling urologic devices: Microbes and antimicrobial management prospect. *Annals of Medical and Health Sciences Research*, *4*(1), 100. https://doi.org/10.4103/2141-9248.126612

Malachowa, N., & Deleo, F. R. (2010). Mobile genetic elements of Staphylococcus aureus. In *Cellular and Molecular Life Sciences* (Vol. 67, Issue 18). https://doi.org/10.1007/s00018-010-0389-4

Maljkovic Berry, I., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., Morton, L., & Jarman, R. G. (2020). Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. *Journal of Infectious Diseases*, *221*(Supplement_3), S292–S307. https://doi.org/10.1093/infdis/jiz286

Mann, E. E., Rice, K. C., Boles, B. R., Endres, J. L., Ranjit, D., Chandramohan, L., Tsang, L. H., Smeltzer, M. S., Horswill, A. R., & Bayles, K. W. (2009). Modulation of eDNA release and degradation affects Staphylococcus aureus biofilm maturation. *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005822

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., & Bryant, S. H. (2015). CDD: NCBI's conserved domain database INTRODUCTION AND STATISTICS ON CONSERVED DOMAIN DATABASE (CDD) COVERAGE. *Nucleic Acids Research*, *43*. https://doi.org/10.1093/nar/gku1221

Mariutti, R. B., Tartaglia, N. R., Seyffert, N., Castro, T. L. de P., Arni, R. K., Azevedo, V. A., le Loir, Y., & Nishifuji, K. (2017). Exfoliative Toxins of Staphylococcus aureus. In *The Rise of Virulence and Antibiotic Resistance in* Staphylococcus aureus. InTech. https://doi.org/10.5772/66528

Martineau, F., Picard, F. J., Roy, P. H., Ouellette, M., & Bergeron, M. G. (1996). Species-specific and ubiquitous DNA-based assays for rapid identification of Staphylococcus epidermidis. *Journal of Clinical Microbiology*, *34*(12), 2888–2893.

Martínez-Botella, G., Loch, J. T., Green, O. M., Kawatkar, S. P., Olivier, N. B., Boriack-Sjodin, P. A., & Keating, T. A. (2013). Sulfonylpiperidines as novel, antibacterial inhibitors of Gram-positive thymidylate kinase (TMK). *Bioorganic & Medicinal Chemistry Letters*, *23*(1). https://doi.org/10.1016/j.bmcl.2012.10.128

Mathur, T., Singhal, S., Khan, S., Upadhyay, D. J., Fatma, T., & Rattan, A. (2006). Detection of biofilm formation among the clinical isolates of Staphylococci: an evaluation of three different screening methods. *Indian journal of medical microbiology*, *24*(1), 25–29. https://doi.org/10.4103/0255-0857.19890

McCarthy, A. J., Breathnach, A. S., & Lindsay, J. A. (2012). Detection of mobile-genetic-element variation between colonizing and infecting hospital-associated methicillin-resistant Staphylococcus aureus isolates. *Journal of Clinical Microbiology*, *50*(3), 1073–1075. https://doi.org/10.1128/JCM.05938-11

McCarthy, A. J., & Lindsay, J. A. (2010). Genetic variation in staphylococcus aureus surface and immune evasion genes is lineage associated: Implications for vaccine design and host-pathogen interactions. *BMC Microbiology*, *10*. https://doi.org/10.1186/1471-2180-10-173

McCarthy, A. J., & Lindsay, J. A. (2012). The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated. *BMC Microbiology*, *12*(1), 104. https://doi.org/10.1186/1471-2180-12-104

McCarthy, A. J., Loeffler, A., Witney, A. A., Gould, K. A., Lloyd, D. H., & Lindsay, J. A. (2014). Extensive horizontal gene transfer during staphylococcus aureus co-colonization in vivo. *Genome Biology and Evolution*, *6*(10), 2697–2708. https://doi.org/10.1093/gbe/evu214

McCarthy, A. J., Witney, A. A., & Lindsay, J. A. (2012). Staphylococcus aureus temperate bacteriophage: carriage and horizontal gene transfer is lineage associated. *Frontiers in Cellular and Infection Microbiology*, *2*, 6. https://doi.org/10.3389/fcimb.2012.00006

McClure, J. A. M., Lakhundi, S., Kashif, A., Conly, J. M., & Zhang, K. (2018). Genomic comparison of highly virulent, moderately virulent, and avirulent strains from a genetically closely-related MRSA ST239 sub-lineage provides insights into pathogenesis. *Frontiers in Microbiology*, *9*(JUL). https://doi.org/10.3389/fmicb.2018.01531

McGuinness, W. A., Malachowa, N., & DeLeo, F. R. (2017). Vancomycin resistance in Staphylococcus aureus. *Yale Journal of Biology and Medicine*, *90*, 269–281.

McNair, K., Bailey, B. A., & Edwards, R. A. (2012). PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics*, *28*(5). https://doi.org/10.1093/bioinformatics/bts014

Mehndiratta, P. L., & Bhalla, P. (2012). Typing of Methicillin resistant Staphylococcus aureus: A technical review. In *Indian Journal of Medical Microbiology* (Vol. 30, Issue 1). https://doi.org/10.4103/0255-0857.93015

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*. https://doi.org/10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10210-3

Melehani, J. H., James, D. B. A., DuMont, A. L., Torres, V. J., & Duncan, J. A. (2015). Staphylococcus aureus Leukocidin A/B (LukAB) Kills Human Monocytes via Host NLRP3 and ASC when Extracellular, but Not Intracellular. *PLoS Pathogens*, *11*(6). https://doi.org/10.1371/journal.ppat.1004970

Melish, M. E., & Glasgow, L. A. (1970). The Staphylococcal Scalded-Skin Syndrome: Development of an Experimental Model. *New England Journal of Medicine*, *282*(20). https://doi.org/10.1056/NEJM197005142822002

Mistry, H., Sharma, P., Mahato, S., Saravanan, R., Kumar, P. A., & Bhandari, V. (2016). Prevalence and characterization of oxacillin susceptible mecA-Positive clinical isolates of

staphylococcus aureus causing bovine mastitis in india. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162256

Młynarczyk, A., Młynarczyk, G., & Jeljaszewicz, J. (1998). The genome of Staphylococcus aureus: A review. *Zentralblatt Fur Bakteriologie*, *287*(4). https://doi.org/10.1016/S0934-8840(98)80165-5

Mondal, S. I., Ferdous, S., Jewel, N. A., Akter, A., Mahmud, Z., Islam, M. M., Afrin, T., & Karim, N. (2015). Identification of potential drug targets by subtractive genome analysis of Escherichia coli O157:H7: An in silico approach. *Advances and Applications in Bioinformatics and Chemistry*, *8*(1), 49–63. https://doi.org/10.2147/AABC.S88522

Monecke, S., Coombs, G., Shore, A. C., Coleman, D. C., Akpaka, P., Borg, M., Chow, H., Ip, M., Jatzwauk, L., Jonas, D., Kadlec, K., Kearns, A., Laurent, F., O'Brien, F. G., Pearson, J., Ruppelt, A., Schwarz, S., Scicluna, E., Slickers, P., … Ehricht, R. (2011). A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant Staphylococcus aureus. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0017936

Monterrubio-López, G. P., González-Y-Merchand, J. A., & Ribas-Aparicio, R. M. (2015). Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/483150

Møretrø, T., Hermansen, L., Holck, A. L., Sidhu, M. S., Rudi, K., & Langsrud, S. (2003). Biofilm Formation and the Presence of the Intercellular Adhesion Locus ica among Staphylococci from Food and Food Processing Environments. *Applied and Environmental Microbiology*, *69*(9), 5648–5655. https://doi.org/10.1128/AEM.69.9.5648-5655.2003

Moroney, S. M., Heller, L. C., Arbuckle, J., Talavera, M., & Widen, R. H. (2007). Staphylococcal Cassette Chromosome mec and Panton-Valentine Leukocidin Characterization of Methicillin-Resistant Staphylococcus aureus Clones. *Journal of Clinical Microbiology*, *45*(3), 1019–1021. https://doi.org/10.1128/JCM.01706-06

Morpheus. (n.d.). Retrieved February 8, 2019, from https://software.broadinstitute.org/morpheus/

Morya, V. K., Dewaker, V., & Kim, E. K. (2012). In Silico Study and Validation of Phosphotransacetylase (PTA) as a Putative Drug Target for Staphylococcus aureus by Homology-Based Modelling and Virtual Screening. *Applied Biochemistry and Biotechnology*, *168*(7). https://doi.org/10.1007/s12010-012-9897-z

Motin, V., & Torres, A. (2009). Molecular Approaches to Bacterial Vaccines. *Vaccines For Biodefense and Emerging And Neglected Diseases*, 63-76. doi: 10.1016/b978-0-12-369408-9.00006-8

Mottola, C., Matias, C. S., Mendes, J. J., Melo-Cristino, J., Tavares, L., Cavaco-Silva, P., & Oliveira, M. (2016). Susceptibility patterns of Staphylococcus aureus biofilms in diabetic foot infections. *BMC Microbiology*, *16*(1). https://doi.org/10.1186/s12866-016-0737-0

Muzzi, A., Masignani, V., & Rappuoli, R. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. In *Drug Discovery Today* (Vol. 12, Issues 11–12, pp. 429–439). Elsevier Current Trends. https://doi.org/10.1016/j.drudis.2007.04.008

Narita, S., Kaneko, J., Chiba, J. ichi, Piémont, Y., Jarraud, S., Etienne, J., & Kamio, Y. (2001). Phage conversion of Panton-Valentine leukocidin in Staphylococcus aureus: Molecular analysis of a PVL-converting phage, φSLT. *Gene*, *268*(1–2). https://doi.org/10.1016/S0378-1119(01)00390-0

Nasr, R. A., AbuShady, H. M., & Hussein, H. S. (2012). Biofilm formation and presence of icaAD gene in clinical isolates of staphylococci. *Egyptian Journal of Medical Human Genetics*, *13*(3). https://doi.org/10.1016/j.ejmhg.2012.04.007

Needham, C., Noble, W. C., & Dyke, K. G. H. (1995). The staphylococcal insertion sequence is257is active. *Plasmid*, *34*(3), 198–205. https://doi.org/10.1006/plas.1995.0005

Nessler, S. (2005). The bacterial HPr kinase/phosphorylase: A new type of Ser/Thr kinase as antimicrobial target. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1754*(1–2), 126–131. https://doi.org/10.1016/j.bbapap.2005.07.042

Nguyen, T. L., & Kim, D. H. (2018). Genome-wide comparison reveals a probiotic strain lactococcus lactis wflu12 isolated from the gastrointestinal tract of olive flounder (paralichthys olivaceus) harboring genes supporting probiotic action. *Marine Drugs*, *16*(5). https://doi.org/10.3390/md16050140

Niemann, H. H., Schubert, W. D., & Heinz, D. W. (2004). Adhesins and invasins of pathogenic bacteria: A structural view. In *Microbes and Infection* (Vol. 6, Issue 1, pp. 101–112). Elsevier Masson SAS. https://doi.org/10.1016/j.micinf.2003.11.001

Nishifuji, K., Sugai, M., & Amagai, M. (2008). Staphylococcal exfoliative toxins: "Molecular scissors" of bacteria that attack the cutaneous defense barrier in mammals. In *Journal of Dermatological Science* (Vol. 49, Issue 1). https://doi.org/10.1016/j.jdermsci.2007.05.007

Noori Goodarzi, N., Bolourchi, N., Fereshteh, S., Soltani Shirazi, A., Pourmand, M. and Badmasti, F., (2021). Investigation of novel putative immunogenic targets against Staphylococcus aureus using a reverse vaccinology strategy. Infection, Genetics and Evolution, 96, p.105149

Noto, M. J., & Archer, G. L. (2006). A subset of Staphylococcus aureus strains harboring staphylococcal cassette chromosome mec (SCCmec) type IV is deficient in CcrAB-mediated SCCmec excision. *Antimicrobial Agents and Chemotherapy*, *50*(8), 2782–2788. https://doi.org/10.1128/AAC.00032-06

Noto, M. J., Kreiswirth, B. N., Monk, A. B., & Archer, G. L. (2008). Gene acquisition at the insertion site for SCCmec, the genomic island conferring methicillin resistance in Staphylococcus aureus. *Journal of Bacteriology*, *190*(4), 1276–1283. https://doi.org/10.1128/JB.01128-07

Novick, R. P. (1989). Staphylococcal Plasmids and their Replication. *Annual Review of Microbiology*, *43*(1). https://doi.org/10.1146/annurev.mi.43.100189.002541

Novick, R. P., Christie, G. E., & Penadés, J. R. (2010). The phage-related chromosomal islands of Gram-positive bacteria. In *Nature Reviews Microbiology* (Vol. 8, Issue 8, pp. 541–551). NIH Public Access. https://doi.org/10.1038/nrmicro2393

Nowrouzian, F. L., Ali, A., Badiou, C., Dauwalder, O., Lina, G., & Josefsson, E. (2015). Impacts of enterotoxin gene cluster-encoded superantigens on local and systemic experimental Staphylococcus aureus infections. *European Journal of Clinical Microbiology & Infectious Diseases*, *34*(7), 1443–1449. https://doi.org/10.1007/s10096-015-2371-4

Nübel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., Žemličková, H., Leblois, R., Wirth, T., Jombart, T., Balloux, F., & Witte, W. (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant staphylococcus aureus. *PLoS Pathogens*, *6*(4), 1–12. https://doi.org/10.1371/journal.ppat.1000855

Nübel, U., Roumagnac, P., Feldkamp, M., Song, J. H., Ko, K. S., Huang, Y. C., Coombs, G., Ip, M., Westh, H., Skov, R., Struelens, M. J., Goering, R. v., Strommenger, B., Weller, A., Witte, W., & Achtman, M. (2008). Frequent emergence and limited geographic dispersal of methicillin- resistant Staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14130–14135. https://doi.org/10.1073/pnas.0804178105

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., Prjibelski, A. D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S. R., Woyke, T., McLean, J. S., Lasken, R., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of Computational Biology*, *20*(10). https://doi.org/10.1089/cmb.2013.0084

O' Neil, J. (2014). Review on Antibiotic resistance. Antimicrobial Resistance : Tackling a crisis for the health and wealth of nations. In *Health and Wealth Nations* (Issue December).

Omar, N. Y., Ali, H. A. S., Harfoush, R. A. H., & El Khayat, E. H. (2014). Molecular typing of methicillin resistant staphylococcus aureus clinical isolates on the basis of protein a and coagulase gene polymorphisms. *International Journal of Microbiology*, *2014*. https://doi.org/10.1155/2014/650328

O'Neill, E., Pozzi, C., Houston, P., Humphreys, H., Robinson, D. A., Loughman, A., Foster, T. J., & O'Gara, J. P. (2008). A novel Staphylococcus aureus biofilm phenotype mediated by the fibronectin-binding proteins, FnBPA and FnBPB. *Journal of Bacteriology*, *190*(11), 3835–3850. https://doi.org/10.1128/JB.00167-08

O'Neill, E., Pozzi, C., Houston, P., Smyth, D., Humphreys, H., Robinson, D. A., & O'Gara, J. P. (2007). Association between methicillin susceptibility and biofilm regulation in Staphylococcus aureus isolates from device-related infections. *Journal of Clinical Microbiology*, *45*(5), 1379–1388. https://doi.org/10.1128/JCM.02280-06

Ong, E., Cooke, M., Huffman, A., Xiang, Z., Wong, M., & Wang, H. et al. (2021). Vaxign2: the second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning. *Nucleic Acids Research*, *49*(W1), W671-W678. doi: 10.1093/nar/gkab279

O'Riordan, K., & Lee, J. C. (2004). Staphylococcus aureus Capsular Polysaccharides. In *Clinical Microbiology Reviews* (Vol. 17, Issue 1, pp. 218–234). American Society for Microbiology (ASM). https://doi.org/10.1128/CMR.17.1.218-234.2004

Otto, M. (2008). Staphylococcal biofilms. In *Current Topics in Microbiology and Immunology* (Vol. 322, pp. 207–228). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75418-3_10

Otto, M. (2010). Basis of virulence in community-associated methicillin-resistant Staphylococcus aureus. In *Annual Review of Microbiology* (Vol. 64, pp. 143–162). Annu Rev Microbiol. https://doi.org/10.1146/annurev.micro.112408.134309

Otto, M. (2012). Methicillin-resistant Staphylococcus aureus infection is associated with increased mortality. *Future Microbiology*, *7*(2), 189–191. https://doi.org/10.2217/fmb.11.156

Otto, M. (2014). Staphylococcus aureus toxins. In *Current Opinion in Microbiology* (Vol. 17, Issue 1, pp. 32–37). Curr Opin Microbiol. https://doi.org/10.1016/j.mib.2013.11.004

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1226

Ozer, E. A. (2018). ClustAGE: A tool for clustering and distribution analysis of bacterial accessory genomic elements. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2154-x

Ozer, E. A., Allen, J. P., & Hauser, A. R. (2014). Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics*, *15*(1), 1–17. https://doi.org/10.1186/1471-2164-15-737

Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M. N., Iimura, S., Gajiwala, K., Scholtz, J. M., & Grimsley, G. R. (2011). Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology*, *408*(3). https://doi.org/10.1016/j.jmb.2011.02.053

Pantosti, A., Sanchini, A., & Monaco, M. (2007). Mechanisms of antibiotic resistance in Staphylococcus aureus. In *Future Microbiology* (Vol. 2, Issue 3, pp. 323–334). Future Microbiol. https://doi.org/10.2217/17460913.2.3.323

Parker, M. W., & Feil, S. C. (2005). Pore-forming protein toxins: From structure to function. In *Progress in Biophysics and Molecular Biology* (Vol. 88, Issue 1, pp. 91–142). Pergamon. https://doi.org/10.1016/j.pbiomolbio.2004.01.009

Peacock, S. J., Day, N. P., Thomas, M. G., Berendt, A. R., & Foster, T. J. (2000). Clinical isolates of Staphylococcus aureus exhibit diversity in fnb genes and adhesion to human fibronectin. *The Journal of Infection*, *41*(1), 23–31. https://doi.org/10.1053/jinf.2000.0657

Peacock, S. J., Foster, T. J., Cameron, B. J., & Berendt, A. R. (1999). Bacterial fibronectin-binding proteins and endothelial cell surface fibronectin mediate adherence of Staphylococcus aureus to resting human endothelial cells. *Microbiology*, *145*(12), 3477–3486. https://doi.org/10.1099/00221287-145-12-3477

Périchon, B., & Courvalin, P. (2009). VanA-type vancomycin-resistant Staphylococcus aureus. In *Antimicrobial Agents and Chemotherapy* (Vol. 53, Issue 11). https://doi.org/10.1128/AAC.00346-09

Perumal, D., Lim, C. S., Sakharkar, K. R., & Sakharkar, M. K. (2007). Differential genome analyses of metabolic enzymes in pseudomonas aeruginosa for drug target identification. *In Silico Biology*, *7*(4–5), 453–465. https://europepmc.org/article/med/18391237

Pinho, M. G., Filipe, S. R., de Lencastre, H., & Tomasz, A. (2001). Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (PBP2) by the drug resistance protein PBP2A in Staphylococcus aureus. *Journal of Bacteriology*, *183*(22), 6525–6531. https://doi.org/10.1128/JB.183.22.6525-6531.2001

Pizarro-Cerdá, J., & Cossart, P. (2006). Bacterial adhesion and entry into host cells. In *Cell* (Vol. 124, Issue 4, pp. 715–727). Elsevier B.V. https://doi.org/10.1016/j.cell.2006.02.012

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Aricò, B., Comanducci, M., Jennings, G. T., Baldi, L., Bartolini, E., Capecchi, B., Galeotti, C. L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., … Rappuoli, R. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*, *287*(5459), 1816–1820. https://doi.org/10.1126/science.287.5459.1816

Pope, W. H. (2019). Genetic Mosaicism in the Tailed dsDNA Phages. In *Reference Module in Life Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20961-8

Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., Jacobs, W. R., Hendrix, R. W., Lawrence, J. G., & Hatfull, G. F. (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *ELife*, *4*. https://doi.org/10.7554/eLife.06416

Pradhan, S., & Sinha, C. (2018). High throughput screening against pantothenate synthetase identifies amide inhibitors against Mycobacterium tuberculosis and Staphylococcus aureus. *In Silico Pharmacology*, *6*(1), 9. https://doi.org/10.1007/s40203-018-0046-4

Price, L. B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P. S., Pearson, T., Waters, A. E., Foster, J. T., Schupp, J., Gillece, J., Driebe, E., Liu, C. M., Springer, B., Zdovc, I., Battisti, A., Franco, A., Żmudzki, J., Schwarz, S., … Aarestrup, F. M. (2012). Staphylococcus aureus CC398: Host Adaptation and Emergence of Methicillin Resistance in Livestock. *MBio*, *3*(1). https://doi.org/10.1128/mBio.00305-11

Pugazhendhi, A., Michael, D., Prakash, D., Priyadarshini Krishnamaurthy, P., Shanmuganathan, R., Abdullah Al-Dhabi, N., Duraipandiyan, V., Valan Arasu, M., & Kaliannan, T. (2020). Antibiogram and plasmid profiling of beta-lactamase producing multi drug resistant Staphylococcus aureus isolated from poultry litter. *Journal of King Saud University - Science*, *32*(6), 2723–2727. https://doi.org/10.1016/j.jksus.2020.06.007

Que, Y. A., & Moreillon, P. (2014). Staphylococcus aureus (Including Staphylococcal Toxic Shock Syndrome). In *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* (Vol. 2). https://doi.org/10.1016/B978-1-4557-4801-3.00196-X

Rahimi, F., Katouli, M., & Karimi, S. (2016). Biofilm production among methicillin resistant Staphylococcus aureus strains isolated from catheterized patients with urinary tract infection. *Microbial Pathogenesis*, *98*, 69–76. https://doi.org/10.1016/j.micpath.2016.06.031

Rappuoli, R. (2000). Reverse vaccinology. In *Current Opinion in Microbiology* (Vol. 3, Issue 5, pp. 445–450). Current Biology Ltd. https://doi.org/10.1016/S1369-5274(00)00119-3

Rappuoli, R. (2001). Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*, *19*(17–19), 2688–2691. https://doi.org/10.1016/S0264-410X(00)00554-5

Raven, K. E., Blane, B., Kumar, N., Leek, D., Bragin, E., Coll, F., Parkhill, J., & Peacock, S. J. (2020). Defining metrics for whole-genome sequence analysis of MRSA in clinical practice. *Microbial Genomics*, *6*(4), e000354. https://doi.org/10.1099/mgen.0.000354

Reddy, K. G., Rao, N. K., & Rama Krishna, P. B. (2010). In silico identification of potential therapeutic targets in Clostridium botulinum by the approach subtractive genomics. In *International Journal of Bioinformatics Research* (Vol. 2, Issue 2). http://db.psort.org/

Reffuveille, F., Josse, J., Vallé, Q., Mongaret, C., & Gangloff, S. C. (2017). Staphylococcus aureus Biofilms and their Impact on the Medical Field. In *The Rise of Virulence and Antibiotic Resistance in* Staphylococcus aureus. https://doi.org/10.5772/66380

Reniere, M. L., & Skaar, E. P. (2008). Staphylococcus aureus haem oxygenases are differentially regulated by iron and haem. *Molecular Microbiology*, *69*(5), 1304–1315. https://doi.org/10.1111/j.1365-2958.2008.06363.x

Resch, G., François, P., Morisset, D., Stojanov, M., Bonetti, E. J., Schrenzel, J., Sakwinska, O., & Moreillon, P. (2013). Human-to-Bovine Jump of Staphylococcus aureus CC8 Is Associated with the Loss of a β-Hemolysin Converting Prophage and the Acquisition of a New Staphylococcal Cassette Chromosome. *PLoS ONE*, *8*(3), e58187. https://doi.org/10.1371/journal.pone.0058187

Ridom SpaServer - SpaTypes. (n.d.). Retrieved February 9, 2019, from https://spa.ridom.de/spatypes.shtml

Rinaudo, C. D., Telford, J. L., Rappuoli, R., & Seib, K. L. (2009). Vaccinology in the genome era. In *Journal of Clinical Investigation* (Vol. 119, Issue 9, pp. 2515–2525). American Society for Clinical Investigation. https://doi.org/10.1172/JCI38330

Rohde, H., Burandt, E. C., Siemssen, N., Frommelt, L., Burdelski, C., Wurster, S., Scherpe, S., Davies, A. P., Harris, L. G., Horstkotte, M. A., Knobloch, J. K. M., Ragunath, C., Kaplan, J. B., & Mack, D. (2007). Polysaccharide intercellular adhesin or protein factors in biofilm accumulation of Staphylococcus epidermidis and Staphylococcus aureus isolated from prosthetic hip and knee joint infections. *Biomaterials*, *28*(9), 1711–1720. https://doi.org/10.1016/j.biomaterials.2006.11.046

Rossi, A. D., Oliveira, P. H. E., Siqueira, D. G., Reis, V. C. C., Dardenne, L. E., & Goliatt, P. V. Z. C. (2020). MHOLline 2.0: Workflow for automatic large-scale modeling and analysis of proteins. *Revista Mundi Engenharia, Tecnologia e Gestão (ISSN: 2525-4782)*, *5*(6). https://doi.org/10.21575/25254782rmetg2020vol5n61325

Rossney, A. S., Shore, A. C., Morgan, P. M., Fitzgibbon, M. M., O'Connell, B., & Coleman, D. C. (2007). The Emergence and Importation of Diverse Genotypes of Methicillin-Resistant Staphylococcus aureus (MRSA) Harboring the Panton-Valentine Leukocidin Gene (pvl) Reveal that pvl Is a Poor Marker for Community-Acquired MRSA Strains in Ireland. *Journal of Clinical Microbiology*, *45*(8), 2554–2563. https://doi.org/10.1128/JCM.00245-07

Rossolini, G. M., Arena, F., Pecile, P., & Pollini, S. (2014). Update on the antibiotic resistance crisis. In *Current Opinion in Pharmacology* (Vol. 18, pp. 56–60). Elsevier Ltd. https://doi.org/10.1016/j.coph.2014.09.006

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. http://www.rstudio.com/

Ruzin, A., Lindsay, J., & Novick, R. P. (2001). Molecular genetics of SaPI1 - A mobile pathogenicity island in Ataphylococcus aureus. *Molecular Microbiology*, *41*(2). https://doi.org/10.1046/j.1365-2958.2001.02488.x

Sabat, A. J., Budimir, A., Nashev, D., Sá-Leão, R., van Dijl, J. M., Laurent, F., Grundmann, H., & Friedrich, A. W. (n.d.). *Overview of molecular typing methods for outbreak detection and epidemiological surveillance*. 1. Retrieved January 21, 2022, from www.eurosurveillance.orghttp://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20380

Sachdeva, G., Kumar, K., Jain, P., & Ramachandran, S. (2005). SPAAN: A software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics*, *21*(4), 483–491. https://doi.org/10.1093/bioinformatics/bti028

Sakinç, T., Michalski, N., Kleine, B., & Gatermann, S. G. (2009). The uropathogenic species Staphylococcus saprophyticus tolerates a high concentration of d-serine. *FEMS Microbiology Letters*, *299*(1), 60–64. https://doi.org/10.1111/j.1574-6968.2009.01731.x

Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research*, *2017*. https://doi.org/10.1155/2017/2680160

Schmitz, F. J., Jones, M. E., Hofmann, B., Hansen, B., Scheuring, S., Lückefahr, M., Fluit, A., Verhoef, J., Hadding, U., Heinz, H. P., & Köhrer, K. (1998). Characterization of grlA, grlB, gyrA, and gyrB mutations in 116 unrelated isolates of Staphylococcus aureus and effects of mutations on ciprofloxacin MIC. Antimicrobial agents and chemotherapy, 42(5), 1249–1252. https://doi.org/10.1128/AAC.42.5.1249

Schmidt, H., & Hensel, M. (2004). Pathogenicity Islands in Bacterial Pathogenesis. In *Clinical Microbiology Reviews* (Vol. 17, Issue 1, pp. 14–56). Clin Microbiol Rev. https://doi.org/10.1128/CMR.17.1.14-56.2004

Schwarzkopf, A., & Karch, H. (1994). Genetic variation in Staphylococcus aureus coagulase genes: potential and limits for use as epidemiological marker. *Journal of Clinical Microbiology*, *32*(10), 2407–2412.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14). https://doi.org/10.1093/bioinformatics/btu153

Shakeri, F., Shojai, A., Golalipour, M., Rahimi Alang, S., Vaez, H., & Ghaemi, E. A. (2010). *Spa* Diversity among MRSA and MSSA Strains of *Staphylococcus aureus* in North of Iran. *International Journal of Microbiology*, *2010*, 351397. https://doi.org/10.1155/2010/351397

Sharma, S., Chaudhry, V., Kumar, S., & Patil, P. B. (2018). Phylogenomic based comparative studies on Indian and American commensal Staphylococcus epidermidis isolates. *Frontiers in Microbiology*, *9*(FEB). https://doi.org/10.3389/fmicb.2018.00333

Shin, K., Yun, Y., Yi, S., Lee, H. G., Cho, J. C., Suh, K. do, Lee, J., & Park, J. (2013). Biofilm-forming ability of Staphylococcus aureus strains isolated from human skin. *Journal of Dermatological Science*, *71*(2), 130–137. https://doi.org/10.1016/j.jdermsci.2013.04.004

Shopsin, B., Gomez, M., Montgomery, S. O., Smith, D. H., Waddington, M., Dodge, D. E., Bost, D. A., Riehman, M., Naidich, S., & Kreiswirth, B. N. (1999). Evaluation of protein A gene polymorphic region DNA sequencing for typing of Staphylococcus aureus strains. *Journal of Clinical Microbiology*, *37*(11). https://doi.org/10.1128/jcm.37.11.3556-3563.1999

Shopsin, B., Gomez, M., Waddington, M., Riehman, M., & Kreiswirth, B. N. (2000). Use of Coagulase Gene Repeat Region Nucleotide Sequences for Typing of Methicillin-Resistant Staphylococcus aureus Strains. *Journal of Clinical Microbiology*, *38*(9), 3453 LP – 3456. https://doi.org/10.1128/JCM.38.9.3453-3456.2000

Shukla, S. K., Pantrangi, M., Stahl, B., Briska, A. M., Stemper, M. E., Wagner, T. K., Zentz, E. B., Callister, S. M., Lovrich, S. D., Henkhaus, J. K., & Dykes, C. W. (2012). Comparative whole-genome mapping to determine Staphylococcus aureus genome size, virulence motifs, and clonality. *Journal of Clinical Microbiology*, *50*(11). https://doi.org/10.1128/JCM.01168-12

Sieradzki, K., Pinho, M. G., & Tomasz, A. (1999). Inactivated pbp4 in highly glycopeptide-resistant laboratory mutants of Staphylococcus aureus. *Journal of Biological Chemistry*, *274*(27), 18942–18946. https://doi.org/10.1074/jbc.274.27.18942

Singh, A., Goering, R. v., Simjee, S., Foley, S. L., & Zervos, M. J. (2006). Application of molecular techniques to the study of hospital infection. In *Clinical Microbiology Reviews* (Vol. 19, Issue 3). https://doi.org/10.1128/CMR.00025-05

Skaar, E. P., Humayun, M., Bae, T., DeBord, K. L., & Schneewind, O. (2004). Iron-source preference of Staphylococcus aureus infections. *Science*, *305*(5690), 1626–1628. https://doi.org/10.1126/science.1099930

Skov, R., Larsen, A. R., Kearns, A., Holmes, M., Teale, C., Edwards, G., & Hill, R. (2013). Phenotypic detection of MECC-MRSA: Cefoxitin is more reliable than Oxacillin. *Journal of Antimicrobial Chemotherapy*, *69*(1), 133–135. https://doi.org/10.1093/jac/dkt341

Smith, D. L., Harris, A. D., Johnson, J. A., Silbergeld, E. K., & Morris, J. G. (2002). Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(9), 6434–6439. https://doi.org/10.1073/pnas.082188899

Smith, T. C. (2015). Livestock-Associated Staphylococcus aureus: The United States Experience. In *PLoS Pathogens* (Vol. 11, Issue 2, p. e1004564). Public Library of Science. https://doi.org/10.1371/journal.ppat.1004564

Solanki, V., Tiwari, M., & Tiwari, V. (2019). Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41496-4

Solanki, V., Tiwari, M., & Tiwari, V. (2021). Subtractive proteomic analysis of antigenic extracellular proteins and design a multi-epitope vaccine against Staphylococcus aureus. *Microbiology and Immunology*, *65*(8), 302–316. https://doi.org/10.1111/1348-0421.12870

Soltan, M. A., Magdy, D., Solyman, S. M., & Hanora, A. (2020). Design of Staphylococcus aureus New Vaccine Candidates with B and T Cell Epitope Mapping, Reverse Vaccinology, and Immunoinformatics. *OMICS A Journal of Integrative Biology*, *24*(4), 195–204. https://doi.org/10.1089/omi.2019.0183

Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H., & Gouaux, J. E. (1996). Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science*, *274*(5294). https://doi.org/10.1126/science.274.5294.1859

spa Type Finder/Identifier. (n.d.). Retrieved February 8, 2019, from http://spatyper.fortinbras.us/

Spaan, A. N., Schiepers, A., de Haas, C. J. C., van Hooijdonk, D. D. J. J., Badiou, C., Contamin, H., Vandenesch, F., Lina, G., Gerard, N. P., Gerard, C., van Kessel, K. P. M., Henry, T., & van Strijp, J. A. G. (2015). Differential Interaction of the Staphylococcal Toxins Panton–Valentine Leukocidin and γ-Hemolysin CB with Human C5a Receptors. *The Journal of Immunology*, *195*(3). https://doi.org/10.4049/jimmunol.1500604

Spaulding, A. R., Salgado-Pabón, W., Kohler, P. L., Horswill, A. R., Leung, D. Y. M., & Schlievert, P. M. (2013). Staphylococcal and streptococcal superantigen exotoxins. *Clinical Microbiology Reviews*, *26*(3). https://doi.org/10.1128/CMR.00104-12

Spoor, L. E., McAdam, P. R., Weinert, L. A., Rambaut, A., Hasman, H., Aarestrup, F. M., Kearns, A. M., Larsen, A. R., Skov, R. L., & Ross Fitzgerald, J. (2013). Livestock origin for a human pandemic clone of community-associated methicillin-resistant Staphylococcus aureus. *MBio*, *4*(4), 356–369. https://doi.org/10.1128/mBio.00356-13

Springer, B., Orendi, U., Much, P., Höger, G., Ruppitsch, W., Krziwanek, K., Metz-Gercek, S., & Mittermayer, H. (2009). Methicillin-resistant Staphylococcus aureus: A new zoonotic agent? In *Wiener Klinische Wochenschrift* (Vol. 121, Issues 3–4, pp. 86–90). Springer. https://doi.org/10.1007/s00508-008-1126-y

Stapleton, P. D., & Taylor, P. W. (2002). Methicillin resistance in Staphylococcus aureus: mechanisms and modulation. In *Science progress* (Vol. 85, Issue Pt 1, pp. 57–72). Sci Prog. https://doi.org/10.3184/003685002783238870

Stefanaki, C., Ieronymaki, A., Matoula, T., Caroni, C., Polythodoraki, E., Chryssou, S.-E., Kontochristopoulos, G., & Antoniou, C. (2017). Six-Year Retrospective Review of Hospital Data on Antimicrobial Resistance Profile of Staphylococcus aureus Isolated from Skin Infections from a Single Institution in Greece. *Antibiotics*, *6*(4), 39. https://doi.org/10.3390/antibiotics6040039

Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, *55*(11), 2324–2337. https://doi.org/10.1021/ACS.JCIM.5B00559

Strandén, A., Frei, R., & Widmer, A. F. (2003). Molecular typing of methicillin-resistant Staphylococcus aureus: Can PCR replace pulsed-field gel electrophoresis? *Journal of Clinical Microbiology*, *41*(7). https://doi.org/10.1128/JCM.41.7.3181-3186.2003

Strommenger, B., Braulke, C., Heuck, D., Schmidt, C., Pasemann, B., Nubel, U., & Witte, W. (2008). spa Typing of Staphylococcus aureus as a Frontline Tool in Epidemiological Typing. *Journal of Clinical Microbiology*, *46*(2), 574–581. https://doi.org/10.1128/JCM.01599-07

Strommenger, B., Kettlitz, C., Werner, G., & Witte, W. (2003). Multiplex PCR assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in Staphylococcus aureus. *Journal of Clinical Microbiology*, *41*(9). https://doi.org/10.1128/JCM.41.9.4089-4094.2003

Sugimoto, S., Iwamoto, T., Takada, K., Okuda, K. I., Tajima, A., Iwase, T., & Mizunoe, Y. (2013). Staphylococcus epidermidis Esp degrades specific proteins associatedwith staphylococcus aureus biofilm formation and host-pathogen interaction. *Journal of Bacteriology*, *195*(8), 1645–1655. https://doi.org/10.1128/JB.01672-12

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, *27*(7), 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

Surewaard, B. G. J., de Haas, C. J. C., Vervoort, F., Rigby, K. M., Deleo, F. R., Otto, M., van Strijp, J. A. G., & Nijland, R. (2013). Staphylococcal alpha-phenol soluble modulins contribute to neutrophil lysis after phagocytosis. *Cellular Microbiology*, *15*(8). https://doi.org/10.1111/cmi.12130

Tahir ul Qamar, M., Ahmad, S., Fatima, I., Ahmad, F., Shahid, F., Naz, A., Abbasi, S. W., Khan, A., Mirza, M. U., Ashfaq, U. A., & Chen, L. L. (2021). Designing multi-epitope vaccine against Staphylococcus aureus by employing subtractive proteomics, reverse vaccinology and immuno-informatics approaches. *Computers in Biology and Medicine*, *132*, 104389. https://doi.org/10.1016/J.COMPBIOMED.2021.104389

Tallent, S. M., Langston, T. B., Moran, R. G., & Christie, G. E. (2007). Transducing particles of Staphylococcus aureus pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. *Journal of Bacteriology*, *189*(20). https://doi.org/10.1128/JB.00738-07

Tang, F., Bossers, A., Harders, F., Lu, C., & Smith, H. (2013). Comparative genomic analysis of twelve Streptococcus suis (pro)phages. *Genomics*, *101*(6), 336–344. https://doi.org/10.1016/j.ygeno.2013.04.005

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*(14), 6614–6624. https://doi.org/10.1093/NAR/GKW569

Tenover, F. C., & Gaynes, R. P. (2000). The epidemiology of Staphylococcus aureus infections. In V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, & J. I. Rood (Eds.), *Gram-positive pathogens* (pp. 414–421). ASM Press.

Tenover, F. C., Tickler, I. A., Le, V. M., Dewell, S., Mendes, R. E., & Goering, R. v. (2019). Updating molecular diagnostics for detecting methicillin- susceptible and methicillin-resistant staphylococcus aureus Isolates in Blood Culture Bottles. *Journal of Clinical Microbiology*, *57*(11). https://doi.org/10.1128/JCM.01195-19

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. In *Current Opinion in Microbiology* (Vol. 11, Issue 5). https://doi.org/10.1016/j.mib.2008.09.006

Thoden, J. B., & Holden, H. M. (2007). The molecular architecture of glucose-1-phosphate uridylyltransferase. *Protein Science*, *16*(3). https://doi.org/10.1110/ps.062626007

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. In *Nature Reviews Microbiology* (Vol. 3, Issue 9). https://doi.org/10.1038/nrmicro1234

Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321. https://doi.org/10.1021/jm051197e

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., Holland, T. L., & Fowler, V. G. (2019). Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. *Nature Reviews Microbiology 2019 17:4*, *17*(4), 203–218. https://doi.org/10.1038/s41579-018-0147-4

Úbeda, C., Maiques, E., Knecht, E., Lasa, Í., Novick, R. P., & Penadés, J. R. (2005). Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Molecular Microbiology*, *56*(3), 836–844. https://doi.org/10.1111/j.1365-2958.2005.04584.x

Uddin, R., Siddiqui, Q. N., Sufian, M., Azam, S. S., & Wadood, A. (2019). Proteome-wide subtractive approach to prioritize a hypothetical protein of XDR-Mycobacterium tuberculosis as potential drug target. *Genes & Genomics*, *41*(11). https://doi.org/10.1007/s13258-019-00857-z

Uddin, R., & Sufian, M. (2016). Core proteomic analysis of unique metabolic pathways of salmonella enterica for the identification of potential drug targets. *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0146796

Udo, E. E. (2013). Community-Acquired Methicillin-Resistant Staphylococcus aureus: The New Face of an Old Foe? *Medical Principles and Practice*, *22*(Suppl 1), 20. https://doi.org/10.1159/000354201

Udo, E. E., & Al-Sweih, N. (2013). Emergence of methicillin-resistant staphylococcus aureus in the maternity Hospital, Kuwait. *Medical Principles and Practice*, *22*(6), 535–539. https://doi.org/10.1159/000350526

Udo, E. E., Pearman, J. W., & Grubb, W. B. (1993). Genetic analysis of community isolates of methicillin-resistant Staphylococcus aureus in Western Australia. *Journal of Hospital Infection*, *25*(2), 97–108. https://doi.org/10.1016/0195-6701(93)90100-E

Unal, C. M., & Steinert, M. (2014). Microbial Peptidyl-Prolyl cis/trans Isomerases (PPIases): Virulence Factors and Potential Alternative Drug Targets. *Microbiology and Molecular Biology Reviews*, *78*(3). https://doi.org/10.1128/mmbr.00015-14

van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., & Kuipers, O. P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, *46*(W1), W278–W281. https://doi.org/10.1093/nar/gky383

van Wamel, W. J. B., Rooijakkers, S. H. M., Ruyken, M., van Kessel, K. P. M., & van Strijp, J. A. G. (2006). The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on β-hemolysin-converting bacteriophages. *Journal of Bacteriology*, *188*(4). https://doi.org/10.1128/JB.188.4.1310-1315.2006

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M. E., & Etienne, J. (2003). Community-acquired methicillin-resistant staphylococcus aureus carrying panton-valentine leukocidin genes: Worldwide emergence. *Emerging Infectious Diseases*, *9*(8). https://doi.org/10.3201/eid0908.030089

Ventola, C. L. (2015). The antibiotic resistance crisis: causes and threats. *P & T Journal*, *40*(4), 277–283. https://doi.org/Article

Verdon, J., Girardin, N., Lacombe, C., Berjeaud, J. M., & Héchard, Y. (2009). δ-hemolysin, an update on a membrane-interacting peptide. In *Peptides* (Vol. 30, Issue 4, pp. 817–823). Elsevier. https://doi.org/10.1016/j.peptides.2008.12.017

Verkaik, N. J., Benard, M., Boelens, H. A., de Vogel, C. P., Nouwen, J. L., Verbrugh, H. A., Melles, D. C., van Belkum, A., & van Wamel, W. J. B. (2011). Immune evasion cluster-positive bacteriophages are highly prevalent among human Staphylococcus aureus strains, but they are not essential in the first stages of nasal colonization. *Clinical Microbiology and Infection*, *17*(3), 343–348. https://doi.org/10.1111/j.1469-0691.2010.03227.x

Viau, R., Hujer, A. M., Hujer, K. M., Bonomo, R. A., & Jump, R. L. P. (2015). Are Staphylococcus intermedius Infections in Humans Cases of Mistaken Identity? A Case Series and Literature Review. *Open Forum Infectious Diseases*, *2*(3). https://doi.org/10.1093/ofid/ofv110

Vilela Rodrigues, T., Jaiswal, A., de Sarom, A., de Castro Oliveira, L., Freire Oliveira, C., & Ghosh, P. et al. (2019). Reverse vaccinology and subtractive genomics reveal new

therapeutic targets against Mycoplasma pneumoniae : a causative agent of pneumonia. *Royal Society Open Science*, *6*(7), 190907. https://doi.org/10.1098/rsos.190907

Vivona, S., Bernante, F., & Filippini, F. (2006). NERVE: New Enhanced Reverse Vaccinology Environment. *BMC Biotechnology*, *6*. https://doi.org/10.1186/1472-6750-6-35

Vogel, V., Falquet, L., Calderon-Copete, S. P., Basset, P., & Blanc, D. S. (2012). Short term evolution of a highly transmissible methicillin-resistant staphylococcus aureus clone (ST228) in a Tertiary care hospital. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0038969

Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, *28*(15). https://doi.org/10.1093/bioinformatics/bts310

von Delft, F., Lewendon, A., Dhanaraj, V., Blundell, T. L., Abell, C., & Smith, A. G. (2001). The crystal structure of E. Coli pantothenate synthetase confirms it as a member of the cytidylyltransferase superfamily. *Structure*, *9*(5), 439–450. https://doi.org/10.1016/S0969-2126(01)00604-9

Vuong, C., Kocianova, S., Voyich, J. M., Yao, Y., Fischer, E. R., DeLeo, F. R., & Otto, M. (2004). A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *Journal of Biological Chemistry*, *279*(52). https://doi.org/10.1074/jbc.M411374200

Waldrop, R., McLaren, A., Calara, F., & McLemore, R. (2014). Biofilm growth has a threshold response to glucose in vitro. Clinical orthopaedics and related research, 472(11), 3305–3310. https://doi.org/10.1007/s11999-014-3538-5

Walsh, C. (2003). Opinion – anti-infectives: Where will new antibiotics come from? *Nature Reviews Microbiology*, *1*(1). https://doi.org/10.1038/nrmicro727

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., MacHi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., … Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1099

Wen, Q. F., Liu, S., Dong, C., Guo, H. X., Gao, Y. Z., & Guo, F. B. (2019). Geptop 2.0: An updated, more precise, and faster Geptop server for identification of prokaryotic essential genes. *Frontiers in Microbiology*, *10*(JUN). https://doi.org/10.3389/fmicb.2019.01236

Xia, G., & Wolz, C. (2014). Phages of Staphylococcus aureus and their impact on host evolution. *Infection, Genetics and Evolution*, *21*, 593–601. https://doi.org/10.1016/j.meegid.2013.04.022

Yao, B., Zhang, L., Liang, S., & Zhang, C. (2012). SVMTriP: A Method to Predict Antigenic Epitopes Using Support Vector Machine to Integrate Tri-Peptide Similarity and Propensity. *PLoS ONE*, *7*(9). https://doi.org/10.1371/JOURNAL.PONE.0045152

Yao, J., & Rock, C. O. (2018). Therapeutic targets in chlamydial fatty acid and phospholipid synthesis. In *Frontiers in Microbiology* (Vol. 9, Issue SEP). https://doi.org/10.3389/fmicb.2018.02291

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), 1608–1615. https://doi.org/10.1093/bioinformatics/btq249

Zarantonelli, M. L., Antignac, A., Lancellotti, M., Guiyoule, A., Alonso, J. M., & Taha, M. K. (2006). Immunogenicity of meningococcal PBP2 during natural infection and protective activity of anti-PBP2 antibodies against meningococcal bacteraemia in mice. *Journal of Antimicrobial Chemotherapy*, *57*(5). https://doi.org/10.1093/jac/dkl066

Zhang, K., McClure, J. A., Elsayed, S., Louie, T., & Conly, J. M. (2005). Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome mec types I to V in methicillin-resistant Staphylococcus aureus. *Journal of Clinical Microbiology*, *43*(10), 5026–5033. https://doi.org/10.1128/JCM.43.10.5026-5033.2005

Zhang, K., McClure, J.-A., & Conly, J. M. (2012). Enhanced multiplex PCR assay for typing of staphylococcal cassette chromosome mec types I to V in methicillin-resistant Staphylococcus aureus. *Molecular and Cellular Probes*, *26*(5), 218–221. https://doi.org/10.1016/j.mcp.2012.04.002

Zhang, Y., Xu, D., Shi, L., Cai, R., Li, C., & Yan, H. (2018). Association between agr type, virulence factors, biofilm formation and antibiotic resistance of Staphylococcus aureus isolates from pork production. *Frontiers in Microbiology*, *9*(JUL). https://doi.org/10.3389/fmicb.2018.01876

Zhang, Z., & Ren, Q. (2015). Why are essential genes essential? - The essentiality of Saccharomyces genes. In *Microbial Cell* (Vol. 2, Issue 8). https://doi.org/10.15698/mic2015.08.218

Zhou, H., Du, H., Zhang, H., Shen, H., Yan, R., He, Y., Wang, M., & Zhu, X. (2013). EsxA might as a virulence factor induce antibodies in patients with Staphylococcus aureus infection. *Brazilian Journal of Microbiology*, *44*(1). https://doi.org/10.1590/S1517-83822013005000019

Zmantar, T., Chaieb, K., Makni, H., Miladi, H., Abdallah, F. Ben, Mahdouani, K., & Bakhrouf, A. (2008). Detection by PCR of adhesins genes and slime production in clinical Staphylococcus aureus. *Journal of Basic Microbiology*, *48*(4), 308–314. https://doi.org/10.1002/jobm.200700289

Acknowledgements

The happiness, and euphoria that come along with the successful completion of any works would be incomplete unless we mentioned the names of the people who made it possible, whose constant guidance and encouragement served as a beam of light and crowned out the efforts.

Taking this opportunity, I would like to thank God, The Almighty who provided me much-needed mental and physical strength for the successful completion of this research work.

This research work would not be possible without the research grant. So, I would like to thank Stipendium Hungaricum Scholarship, Hungary, and University Grant Commission, India. I owe my sincere thanks to the Faculty of Sciences, and Szentágothai Research Center, University of Pecs, Pecs, Hungary for providing me a platform to conduct my research.

I would like to express my profound gratitude towards my supervisor Dr. Csaba Fekete, Ph.D., habil for his valuable guidance, moral and financial supports, and blessing rendered throughout my study.

I would like to thank all my colleagues of the Department of General and Environmental Microbiology, especially Peter Urban, for his indispensable technical assistance and scientific works in the lab. With no words, I would express my deep gratitude to all my friends, and Dr. Gunajit Goswami who always believed in me and stood with me in good and bad times. Also, providing me motivating guidance and a tremendous help to my research career. I am thankful to my wife, Bandana Leishangthem, and my junior Bandana Pangabam who have contributed directly or indirectly during my Ph.D. research.

I heartily thankful to my parents, and for their moral support filled with affection, tolerance, concern, and encouragement throughout my study.

List of publications

Publications related to thesis topic:

1. **Naorem, R. S**., Urban, P., Goswami, G., & Fekete, C. (2020). Characterization of methicillin-resistant *Staphylococcus aureus* through genomics approach. *3 Biotech*, *10*(9), 401. doi:10.1007/s13205-020-02387-y **(Q2; IF: 2.406; 2020)**
2. **Naorem, R. S.**, Blom, J., & Fekete, C. (2021). Genome-wide comparison of four MRSA clinical isolates from Germany and Hungary. *PeerJ, 9*, e10185. doi:10.7717/peerj.10185 **(Q1; IF: 2.98; 2020)**
3. **Naorem, R. S.**, Goswami, G., Gyorgy, S., & Fekete, C. (2021). Comparative analysis of prophages carried by human and animal-associated *Staphylococcus aureus* strains spreading across the European regions. *Scientific reports 11*, 18994. https://doi.org/10.1038/s41598-021-98432-8 **(Q1; IF: 4.379; 2020)**
4. **Naorem, R. S.,** Pangabam, B., Bora, S., Goswami, G., Barooah, M., Hazarika, D., & Fekete, C. (2022). Identification of Putative Vaccine and Drug Targets against the Methicillin-Resistant *Staphylococcus aureus* by Reverse Vaccinology and Subtractive Genomics Approaches. *Molecules*, 27(7), 2083. https://doi.org/10.3390/molecules27072083 **(Q1; IF: 4.411; 2020)**

Publications outside to thesis topic:

1. Bouchelaghem, S., Das, S., **Naorem, R. S**., Czuni, L., Papp, G., & Kocsis, M. (2022). Evaluation of Total Phenolic and Flavonoid Contents, Antibacterial and Antibiofilm Activities of Hungarian Propolis Ethanolic Extract against *Staphylococcus aureus*. *Molecules, 27(2)*, 574. http://dx.doi.org/10.3390/molecules27020574.

(**Q1; IF: 4.411; 2020**)

Conference proceedings related to thesis topic:

1. **Naorem, R. S.,** Urban, P., & Fekete, C. (2019). Comparative analysis of (pro)phages carried by human and livestock-associated *Staphylococcus aureus* strains spreading across the European regions. *In: XVI. János Szentágothai Multidisciplinary Conference and Student Competition*, p. 251, ISBN: 9789634293446
2. **Naorem, R. S**., Urban, P., Czuni, L., Gazdag, Z., Papp, G., & Fekete, C. (2018). Genome-based comparative analysis of Methicillin-resistant *Staphylococcus aureus* strains. *In: Bódog, Ferenc (eds.). 7th Interdisciplinary Doctoral Conference*, 2018, p. 183, ISBN: 9789634292104

Conference proceedings outside to thesis topic:

1. Bouchelaghem, S., Das, S., Czuni, L., Gazdag, Z., Fekete, C., Kőszegi, T., **Naorem, R, S.,** Papp, G. (2019). Antimicrobial activities of Hungarian propolis alone and in combination with antibiotics and its antibiofilm activity on *Staphylococcus aureus*. *In: XVI. János Szentágothai Multidisciplinary Conference and Student Competition,* p. 16, ISBN: 9789634293446
2. Urbán, P., Kovács-Valasek, A., Schönhardt, K., Czuni, L., **Singh Romen N**, Papp, G., Gazdag, Z., Manczinger, L., Kredics, L., Fekete, C. (2017). NGS based peptaibol synthase prediction in the green mould disease causing *Trichoderma pleuroti*. *In: VI. Magyar Mikológiai Konferencia*, 56(1), 155-157.

Statement

I, the undersigned **Naorem Romen Singh** (Birth name: Naorem Romen Singh; Mother's maiden name: Potshangbam Bheigyabati Devi; Place, and Date of Birth: Imphal, India and 01-03-1990).

Today, I submitted my PhD thesis entitled: **"Deciphering the genetic architecture of methicillin-resistant *Staphylococcus aureus* clinical isolates using whole-genome sequencing"** to Doctoral School of Biology and Sportbiology.

Name of supervisor: Csaba Fekete, Ph.D., habil, Department of General and Environmental Microbiology, University of Pécs.

In addition, I have declared that:

- my doctoral dissertation submitted in the present case were not submitted to other doctoral school (not domestic nor abroad university) earlier,
- my application to doctoral process is not rejected within two years,
- I had no unsuccessful doctoral procedure the last two years,
- my doctorate degree has not cancelled within five years,
- my thesis is an independent work, I have not presented another person's intellectual property as it would be mine, the references are clear and complete, the false or falsified data are not used in my dissertation.

Date:.....

……

Signature of candidate

Supplementary materials

Table S1. Antibiotic susceptibility patterns of *S. aureus* isolates

Strains ID	Origin	Diameter of inhibition zone (mm)		
		Ox-1µg	**Cfox-30µg**	**Ery-15µg**
SA G1	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA G2	HA	11.5±0.7 (R)	6±0 (R)	6±0 (R)
SA G3	HA	28±0 (S)	18.5±0.7 (R)	27±0 (S)
SA G4	HA	6±0 (R)	6±0 (R)	24.5±7.7 (S)
SA G5	HA	6±0 (R)	6±0 (R)	24.5±7.7 (S)
SA G6	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA G7	HA	24.5±0.7 (S)	18±0 (R)	30±0 (S)
SA G8	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA G9	CA	6±0 (R)	6±0 (R)	24±8 (S)
SA G10	CA	6±0 (R)	6±0 (R)	6±0 (R)
SA G11	CA	6±0 (R)	6±0 (R)	10±0 (S)
SA G12	CA	6±0 (R)	7.5±2.1 (R)	30±0 (S)
SA G13	CA	6±0 (R)	6±0 (R)	30±0 (S)
SA G14	CA	6±0 (R)	6±0 (R)	30±0 (S)
SA H15	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H16	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H17	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H18	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H19	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H20	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H21	HA	6±0 (R)	6±0 (R)	11±0 (R)
SA H22	HA	6±0 (R)	6±0 (R)	11±0 (R)
SA H23	HA	6±0 (R)	7.5±2.1 (R)	6±0 (R)
SA H24	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H25	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H26	HA	8±2.8 (R)	14.5±0.7 (R)	6±0 (R)
SA H27	HA	6±0 (R)	6±0 (R)	16.14±0 (R)
SA H28	HA	6±0 (R)	6±0 (R)	24.5±0.7 (R)
SA H29	HA	10.5±0.7 (R)	8±2.8 (R)	6±0 (R)
SA H30	HA	6±0 (R)	3±0 (R)	6±0 (R)
SA H31	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H32	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H33	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H34	HA	6±0 (R)	6±0 (R)	6±0 (R)
SA H35	HA	6±0 (R)	6±0 (R)	6±0 (R)
ATCC700698	-	6±0 (R)	6±0 (R)	6±0 (R)
ATCC25923	-	26.5±0.7 (S)	25±0 (S)	33.5±0.7 (S)

[a]**SA G** and **SA H** represent strains isolated from Germany and Hungary**. HA** and **CA** represent the Hospital and Community-associated *S. aureus* strains. Zone of inhibition in mm is given as **Mean ± SD; R** and **S** denote the resistance and susceptible. **Ox,** Oxacillin (≤10mm=Resistant); **Cfox**, Cefoxitin (≤21mm=Resistant); **Ery**, Erythromycin (≤13mm=Resistant) based on CLSI 2014 M100-S24.

Table S2. Biochemical tests and PCR based molecular detection of *mecA*, *pvl*, and *SCCmec* genes of *S. aureus* isolates

[a]Strains ID	Biochemical tests								Molecular detection		
	Cat[a]	Coa[b]	DNase	Citrate	Ure	MF[c]	BL[d]	BF	*mecA*	*pvl*	*SCCmec* type
SA G1	+	+	+	+	+	+	+	-	+	-	I
SA G2	+	+	+	+	+	+	+	-	+	-	I
SA G3	+	+	+	+	+	+	+	+	-	-	NT
SA G4	+	+	+	+	+	+	+	+	+	-	NT
SA G5	+	+	+	+	+	+	+	+	+	-	IVd
SA G6	+	+	+	+	+	+	+	-	+	-	I
SA G7	+	+	+	+	+	+	+	+	-	-	NT
SA G8	+	+	+	+	+	+	+	+	+	+	II
SA G9	+	+	+	+	+	+	+	+	+	+	IVa
SA G10	+	+	+	+	+	+	+	+	+	+	IVa
SA G11	+	+	+	+	+	+	+	+	+	+	IVa
SA G12	+	+	+	+	+	+	+	+	+	-	IVa
SA G13	+	+	+	+	+	+	+	+	+	-	IVa
SA G14	+	+	+	+	+	+	+	+	+	+	V
SA H15	+	+	+	+	+	+	+	+	+	-	II
SA H16	+	+	+	+	+	+	+	-	+	+	II
SA H17	+	+	+	+	+	+	+	+	+	-	II
SA H18	+	+	+	+	+	+	+	+	+	-	I
SA H19	+	+	+	+	+	+	+	+	+	-	I
SA H20	+	+	+	+	+	+	+	+	+	-	II
SA H21	+	+	+	+	+	+	+	+	+	-	IVb
SA H22	+	+	+	+	+	+	+	+	+	-	II
SA H23	+	+	+	+	+	+	+	+	+	-	IVa
SA H24	+	+	+	+	+	+	+	+	+	-	II
SA H25	+	+	+	+	+	+	+	+	+	-	II
SA H26	+	+	+	+	+	+	+	+	+	-	I
SA H27	+	+	+	+	+	+	+	+	+	-	IVd
SA H28	+	+	+	+	+	+	+	+	+	-	IVa
SA H29	+	+	+	+	+	+	+	+	+	-	IVd
SA H30	+	+	+	+	+	+	+	+	+	+	II
SA H31	+	+	+	+	+	+	+	+	+	-	I
SA H32	+	+	+	+	+	+	+	+	+	-	NT
SA H33	+	+	+	+	+	+	+	+	+	-	II
SA H34	+	+	+	+	+	+	+	+	+	+	II
SA H35	+	+	+	+	+	+	+	+	+	-	NT
ATCC700698	+	+	+	+	+	+	+	+	+	-	II
ATCC25923	+	+	+	+	+	+	+	+	-	-	-

Cat[a] , Coa[b], MF[c], BL[d], and **BF** represent catalase, coagulase, mannitol productions, represents hemolysis on blood agar, and biofilm formation, respectively**, *mecA*** represents gene encode for penicillin-binding protein 2a (PBP2a); ***pvl*** represents gene encode for Panton-Valentine leukocidin toxin. + and – represent present and absent; **NT** denotes *SCCmec* cassette non-typeable.

Table S3. Primers used for the detection of biofilm-associated genes

Gene name	Primer Sequence (5′- 3′)	Amplicon Size (bp)
icaA	TTTCGGGTGTCTTCACTCTATTT	141
	TGGCAAGCGGTTCATACTT	
icaB	ACCGGCAACTGGGTTTATT	137
	GCAAATCGTGGGTATGTGTTTC	
icaC	GCGTTAGCAAATGGAGACTATTG	100
	GCGTGCAAATACCCAAGATAAC	
icaD	AGCCCAGACAGAGGGAATA	78
	ACGATATAGCGATAAGTGCTGTC	
icaR	GCTGTTTCTTGAAAGTTGGTATTTG	102
	AGTAGCGAATACACTTCATCTTTG	
fnbA	AACAATCTTAGGTACGGCATTAGA	89
	TCTTGTCCCATCCCAACAAC	
fnbB	GTAGAGGAAAGTGGGAGTTCAG	105
	TGTCGCGCTGTATGATTGT	
ebpS	GTGGCATGGCCAAAGTATTG	77
	CATGCCTCCAAATATCGCTAATG	
clfA	CACAACAGGAAACGACACAATC	115
	TGAGTTGTTGCCGGTGTATTA	
clfB	CACAAACAGTGCGAATGTAGATAG	77
	CTGGCTCTGTTGTAGTGGTATT	
cna	CAGGTGGGTCAAGCAGTTATTA	91
	CTGCAAATCCCGAAACATCAC	

Table S4. Quantification of *S. aureus* biofilm-forming ability

Strain ID	Absorbance at 540nm	Category
SA G1	0.088±0.012	Non
SA G2	0.092±0.015	Non
SA G6	0.085±0.018	Non
SA H16	0.066±0.016	Non
SA G8	0.128±0.032	Moderate
SA G9	0.160±0.009	Moderate
SA G10	0.157±0.062	Moderate
SA H19	0.142±0.011	Moderate
SA H26	0.122±0.037	Moderate
SA H32	0.133±0.051	Moderate
SA G5	0.604±0.10	Strong
SA G11	0.442±0.013	Strong
SA H22	0.490±0.150	Strong
SA H27	0.690±0.066	Strong
SA H29	0.684±0.075	Strong
ATCC25923	0.272±0.039	Moderate

Table S5. Features of 16 whole-genome sequences of *S. aureus* strains used for drugs targets and vaccine candidates' identification

Strain ID	ST	CC	Size (bp)	GC%	CDS	Accession No.
SA G5	ST45	CC45	2760385	32.77	2689	CP032160
SA G6	ST228	CC5	2856214	32.79	2734	RAHA00000000
SA G8	ST225	CC8	2857863	32.81	2743	QZFC00000000
SA H27	ST22	CC22	2783185	32.73	2630	CP032161
SA H29	ST8	CC8	2834624	32.65	2843	CP032468
SA H32	ST22	CC22	2786627	32.72	2657	RAHP00000000
NCTC 8325*	ST8	CC8	2821361	32.9	2872	CP000253.1
CA-347*	ST45	CC45	2875156	32.9	2766	CP006044.1
ST228*	ST228	CC5	2783086	32.82	2654	HE579071.1
JH9*	ST105	CC5	2937129	32.9	2879	CP000703.1
Newman*	ST254	CC8	2878897	32.89	2854	AP009351.1
HO 5096 0412*	ST22	CC22	2832299	32.8	2527	HE681097.1
Mu50 DNA*	ST5	CC5	2878529	32.87	2867	BA000017.4
MRSA252*	ST36	CC30	2902619	32.80	2872	BX571856.1
H-EMRSA-15*	ST22	CC22	2846320	32.80	2740	CP007659
DSM 20231*	ST8	CC8	2755072	32.88	2734	CP011526.1

***** indicates the MRSA strains used as reference; **ST** denotes the MLST type; **CC** represent tht Clonal Complex; and **CDS** represent the coding sequences

Table S6. Typing of *coa* gene and *Hae*III RFLP patterns of *S. aureus* strains

Strain ID	*coa* amplicon size (bp)	RFLP pattern (approx bp)	Pattern code	DI of *coa* PCR typing	*DI of *coa*-PCR-RFLP typing
SA H19	550	400+150	C1	0.8381	0.9619
SA H26	550	400+150	C1		
SA G6	550	405+145	C2		
SA H29	550	200+150+100+90	C3		
SA G11	600	250+200+150	D		
SA G1	650	220+190+150+90	A		
SA G2	650	220+190+150+90	A		
SA G9	660	505+155	E		
SA G5	700	230+200+170+90	B1		
SA H16	700	600+150	B2		
SA H22	700	600+150	B2		
SA H27	700	450+150+100	B3		
SA H32	700	450+150+100	B3		
SA G8	740	590+150	F		
SA G10	800	250+200+150+100	G		

***DI-**Discriminatory index

Table S7. Typing of *spa* gene polymorphism and distribution of different repeats and types of *S. aureus* strains

Strain ID	*spa* amplicon size (bp)	Repeats Succession	Kreiswirth IDs	*spa*-type	*DI of *spa*-PCR typing
SA G1	440	11-19-21-12-21-17-34-24-34-22-25	YHFGFMBQBLO	t051	0.9429
SA G2	500	21-17-34-24-34-22-24-34-22-33-25	FMBQBLQBLPO	UK	
SA G5	480	08-39-34	XE3B	t1011	
SA G6	440	26-17-16	TMK	t535	
SA G8	400	26-17-20-17-12-17-17-16	TMDMGMMK	t003	
SA G9	460	07-23-21-16-16-33-21-16-33-13	UJFKKPFKPE	t175	
SA G10	380	04-20-17-20-17-25-34	ZDMDMOB	t437	
SA G11	460	11-19-12-21-17-34-24-34-22-25	YHGFMBQBLO	t008	
SA H16	355	26-23-17-12-17-16	TJMGMK	t062	
SA H19	355	26-23-17-12-17-16	TJMGMK	t062	
SA H22	460	26-23-17-34-17-82-17-12-17-16	TJMBM[r82]MGMK	t2164	
SA H26	560	26	T1	t458	
SA H27	550	26-23-23-13-23-31-29-17-25-17-25-16-28	TJJEJNF2MOMOKR	t379	
SA H29	430	11-19-12-21-17-34-24-34-22-25	YHGFMBQBLO	t008	
SA H32	550	26-23-23-13-23	TJJEJ	t1258	

***DI**-Discriminatory index calculated without Unknown type; **UK**- Unknown

Table S8. Comparative analysis of VFGs associated with putative prophages

Strains	Size (kb)	ORF	Gene	GC%	Completeness / PHASTER Score	Region Position	Virulence & resistance genes
SA G5	19.8	14.6	29	31.06	Incomplete (20)	841711-861552	*sel, sea, sec, ear*
SA G5	14.6	57.6	24	29.87	Incomplete (50)	1311575-1326199	-
SA G5	57.6	36.3	74	33.64	Intact (100)	1641035-1698669	-
SA G6	36.3	16.4	43	32.0	Questionable (81)	764423-800744	*clfA*
SA G6	16.4	72.8	26	33.4	Incomplete (20)	1046367-1062805	*isdD*
SA G6	72.8	74.5	99	34.6	Intact (150)	1930919-2003736	*sea, sep, sak, scn*, *atl*

SA G6	74.5	8.9	92	33.3	Questionable (70)	2753517-2828049	*lukF-PV*, *lukM, b-lactamase*, *plc*, *aad*A, *aph*A1, *sta*
SA G8	8.9	51.4	17	35.1	Incomplete (50)	250716-259621	-
SA G8	51.4	77.1	87	34.1	Intact (150	659680-711162	*ear*
SA G8	77.1	54.4	95	32.6	Intact (108)	1480513-1557685	*virE, ebp*
SA G8	54.4	32.7	71	32.3	Intact (115)	2002216-2056682	*lukF-PV*, *lukS-PV, lukM*, *plc, sep, sak*, *chp*, *scn*
SA G8	32.7	14.6	39	36.7	Questionable (70)	2827472-2860207	*b-lactamase*
SA H27	42.7	222	63	34.4	Questionable (80)	899479-942221	-
SA H27	45	241	62	31.1	Intact (100)	1741210-1786280	*plc*, *lukF-PV*, *lukM*, *scn*, *sak, sep, sea*
SA H27	16.3	64	21	30.7	Incomplete (30)	1792129-1808491	*ear, seb*
SA H29	71.7	320	88	32.7	Intact (110)	1496964-1568760	*vir*E
SA H29	45.4	235	64	33.0	Intact (110)	2005223-2050666	*plc*, *hlgB, lukF-PV*, *lukM*, *scn*, *chp*, *sak, hlb*
SA H29	16.5	91	19	35.0	Intact (140)	2768835-2785368	*cna*, *atl*
SA H32	104.6	364	131	32.7	Intact (150)	1956029-2060686	*lukF-PV*, *lukM*, *hlgB*, *plc*

Table S9. The list of 60 genomes of *S. aureus* used for analysis of prophages diversity

Strains	Size (bp)	Host Name	Isolation Country	GenBank Accessions	Isolation Source
EDCC5464	2762859	Human	Germany	CP022291	Bone
08-01059	2781170	Human	Germany	JJEX00000000	Nose
09-00736	2780640	Human	Germany	JJEM00000000	Nose
HD1410	2847707	Human	Germany	NXFH01000000	Nose
SA G5	2761186	Human	Germany	CP032160	Infection
SA G6	2854204	Human	Germany	RAHA000000	Skin
SA G8	2863424	Human	Germany	QZFC00000000	Infection
SA H27	2832324	Human	Hungary	CP032468-CP032470	Nostril
SA H29	2785632	Human	Hungary	CP032161	Skin
SA H32	2790120	Human	Hungary	RAHP00000000	Trachea
1111205429	2793560	Human	Netherlands	JJDO00000000	Tracheal aspirate
VET1914R	2769360	Human	Netherlands	JIFV00000000	Nasal
VET1913R	2754890	Human	Netherlands	JIFW00000000	Throat
VET1912R	2765090	Human	Netherlands	JIFX00000000	Nasal
VET1911R	2758890	Human	Netherlands	JIFY00000000	Throat
HL463	2726155	Human	Denmark	OFWS01000000	-
81070.00	2813800	Human	Denmark	JJAI00000000	Blood
81629	2807060	Human	Denmark	JJAH00000000	Wound
I3	2815948	Human	Denmark	MVFW00000000	Knee infection
CM101	2791324	Human	Italy	PZVM01000000	Broncho aspiration material
CM124	2846513	Human	Italy	PZUR01000000	Broncho aspiration material
CM117	2793506	Human	Italy	PZUX01000000	Nasal swab
CM112	2782482	Human	Italy	PZVC01000000	Nasal swab
CM60	2909394	Human	Italy	PZWT01000000	Nose lavage
H4	2888632	Human	Austria	NKCT00000000	-
H2	2890024	Human	Austria	NKCV00000000	-
H1	2870072	Human	Austria	NKCW00000000	-
H8	2875502	Human	Austria	NKCP00000000	-
H6	2719985	Human	Austria	NKCR00000000	-
Ab333	2833925	Human	France	LSGN01000000	Hand skin
P333	2834244	Human	France	LSGO01000000	Nares
SA13-192	2714681	Human	France	LNJF00000000	Infection
SA14-639	2707899	Human	France	LNJO00000000	Infection
SA14-613	2730273	Human	France	LNJN00000000	Infection
Sa52	2762421	Cow	Denmark	NBNF00000000	Udder
CFSAN018750	2719423	Cow	Denmark	CP028189	Infection
22825	2741460	Cow	France	JJBW00000000	Nose
22835	2747140	Cow	France	JJBV00000000	Nose
22837	2742960	Cow	France	JJBU00000000	Nose
22838	2744840	Cow	France	JJBT00000000	Nose
22841	2741720	Cow	France	JJBS00000000	Nose

12-ST01564	2747540	Cow	Germany	JJCW00000000	Nose
Rd.3	2756990	Cow	Germany	JIWZ00000000	Bodily fluid
Rd.9	2877450	Cow	Germany	JIWT00000000	Bodily fluid
Rd.60	2777200	Cow	Germany	JIWV00000000	Bodily fluid
Rd.545	2873820	Cow	Germany	JIWW00000000	Bodily fluid
53180-1	2806390	Cow	Italy	JJAW00000000	Bodily fluid
56824-21	2804880	Cow	Italy	JJAR00000000	Bodily fluid
685	2642470	Cow	Italy	LRNB00000000	Milk
682	2604095	Cow	Italy	LRNA00000000	Milk
909	2697903	Cow	Netherlands	QFCZ01000000	Milk
483	2695752	Cow	Netherlands	QFCY01000000	Milk
FP_N239	2757820	Cow	Netherlands	JIYQ00000000	Nose
FP_N5203 OX	2801580	Cow	Netherlands	JIYP00000000	Nose
FP_N5208 OX	2796570	Cow	Netherlands	JIYO00000000	Nose
2010-60-6511-10	2806950	Cow	Netherlands	JJCG00000000	Bodily fluid
19571	2770430	Dog	France	JJCM00000000	Vagina
23237	2737320	Dog	France	JJBO00000000	Skin
C3489	2765820	Dog	Spain	JIZE00000000	Bodily fluid
C5086	2709270	Dog	Spain	JIYX00000000	Bodily fluid

Table S10. General features of 65 intact prophages extracted from the genomes of *S. aureus* strains

Prophages	Size (Kb)	GC%	CDS	Located region	Score	Host	Origin	Country
phi08-01059-1	46.9	33.23	65	11485-58400	Intact (130)	*S. aureus* 08-01059	Human	Germany
phi09-00736-1	46.9	33.23	66	11416-58331	Intact (130)	*S. aureus* 09-00736	Human	Germany
phiHD1410-1	33.8	32.5	32	287318-321167	Intact (95)	*S. aureus* strain HD1410	Human	Germany: Tubingen
phiHD1410-2	59.7	33.5	65	17033-76828	Intact (120)			
phiHD1410-3	44.9	32.9	64	56719-101698	Intact (100)			
phi1111205429-1	41.2	33.83	56	3-41249	Intact (130)	*S. aureus* 1111205429	Human	Netherlands
phiVET1915R-3	48	33.69	73	243-48313	Intact (140)	*S. aureus* VET1915R	Human	Netherlands
phiVET1914R-1	47.3	33.67	67	137706-185045	Intact (130)	*S. aureus* VET1914R	Human	Netherlands
phiVET1913R-2	47.3	33.67	66	239198-286537	Intact (130)	*S. aureus* strain VET1913R	Human	Netherlands
phiVET1912R-2	43.2	34.86	72	30927-74209	Intact (120)	*S. aureus* strain VET1912R	Human	Netherlands
phi81629-1	30.4	33.78	35	3-30465	Intact (100)	*S. aureus* 81629	Human	Denmark
phiI3-1	33.8	32.5	31	302540-336406	Intact (95)	*S. aureus* strain I3	Human	Denmark: Aalborg
phiI3-2	41.7	33.4	60	1-41759	Intact (108)			
phiI3-4	63.3	32.8	65	33452-96810	Intact (110)			
phiCM101-2	56.7	32.2	68	15026-71738	Intact (130)	*S. aureus* strain CM101	Human	Italy: Florence
phiCM124-1	64.8	33.8	64	21328-86146	Intact (150)	*S. aureus* strain CM124	Human	Italy: Florence
phiCM124-2	60.2	32.5	78	32094-92386	Intact (140)			
phiCM117-1	49.5	34.2	63	153991-203576	Intact (140)	*S. aureus* strain CM117	Human	Italy: Florence
phiCM117-3	46.3	32.8	65	10397-56750	Intact (100)			
phiCM112-3	28.1	34.4	33	729-28851	Intact (110)	*S. aureus* strain CM112	Human	Italy: Florence
phiCM60-3	28.6	35.1	37	257-28905	Intact (100)	*S. aureus* strain CM60	Human	Italy: Florence
phiH4-2	62.2	32.6	73	187368-249650	Intact (140)	*S. aureus* strain H4	Human	Austria: Carinthia
phiH4-3	56.6	32	67	3980-60610	Intact (100)			
phiH2-2	63.2	32.5	73	187298-250551	Intact (140)	*S. aureus* strain H2	Human	Austria: Carinthia
phiH2-3	59.9	32.2	74	31830-91747	Intact (110)			

phiH8-1	44.2	33.2	55	17487-61726	Intact (120)	*S. aureus* strain H8	Human	Austria: Carinthia
phiH3-1	59.9	32.4	70	1136-61074	Intact (110)	*S. aureus* strain H3	Human	Austria: Carinthia
phiH3-2	53.5	33	63	3382-56909	Intact (120)			
phiH1-1	63.2	32.5	73	187368-250621	Intact (140)	*S. aureus* strain H1	Human	Austria: Carinthia
phiH1-3	59.9	32.2	74	31829-91746	Intact (110)			
phiH7-2	24.8	34.4	29	3-24873	Intact (100)	*S. aureus* strain H7	Human	Austria: Vienna
phiAb333-3	60.4	32.6	73	1536165-1596570	Intact (140)	*S. aureus* strainAb333	Human	France: Rennes
phiAb333-4	45.8	32.8	66	162378-208180	Intact (100)			
phiP333-3	60.4	32.6	73	1534491-1594896	Intact (140)	*S. aureus* strain p333	Human	France: Rennes
phiP333-4	45.8	32.8	66	88453-134255	Intact (100)			
phiSA13-192-1	32.1	34	54	524-32652	Intact (110)	*S. aureus* subsp. *aureus* strain SA13-192	Human	France: Tours
phiSA13-192-2	33.8	33.5	53	515-34393	Intact (110)			
phiSA14-639-2	44.5	33	65	13727-58297	Intact (120)	*S. aureus* subsp. *aureus* strain SA14-639	Human	France: Tours
phiG4-3	87.8	33.92	132	1944955-2032780	Intact (150)	*S. aureus* subsp. *aureus* strain SA G6	Human	Germany
phiG5-2	61.2	33.59	90	682931-744225	Intact (150)	*S. aureus* subsp. *aureus* strain SA G8	Human	Germany
phiG5-3	68.2	32.59	79	1517828-1586108	Intact (130)			
phiG5-5	51.8	32.22	59	2040204-2092008	Intact (106)			
phiH13-2	56.4	32.16	66	1013609-1070087	Intact (100)	*S. aureus* subsp. *aureus* strain SA H27	Human	Hungary
phiH14-1	73.3	32.75	73	1502314-1575621	Intact (120)	*S. aureus* subsp. *aureus* strain SA H29	Human	Hungary
phiH14-2	45.5	32.94	64	2018439-2063982	Intact (100)			
phiH15-2	58.7	33.56	70	736189-794911	Intact (150)	*S. aureus* subsp. *aureus* strain SA H32	Human	Germany
phi2010-60-6511-10-1	46.7	34.76	68	222914-269656	Intact (120)	*S. aureus* 2010-60-6511-10	Animal	Netherlands

phi22835-2	44.7	33	68	98989-143737	Intact (100)	*S. aureus* 22835-2	Animal	France
phi53180-1-4	31.5	33.72	36	82-31604	Intact (110)	*S. aureus* 53180-1	Animal	Italy
phiFP_N239-1	46.4	33.55	64	142176-188632	Intact (110)	*S. aureus* FP_N239	Animal	Netherlands
phiFP_N5203 OX-1	63.7	33.15	71	1137-64873	Intact (140)	*S. aureus* FP_N5203 OX	Animal	Netherlands
phiFP_N5208 OX-2	48.1	33.49	62	239490-287619	Intact (120)	*S. aureus* FP_N5208 OX-2	Animal	Netherlands
phi483-4	55.1	33	80	35402-90527	Intact (150)	*S. aureus* 483-4	Animal	Netherlands
phi909-4	55	33.1	80	35713-90759	Intact (140)	*S. aureus* 909-4	Animal	Netherlands
phi22825-1	44.7	33	68	63597-108345	Intact (100)	*S. aureus* 22825	Animal	France
phi22837-1	44.7	33	68	91414-136162	Intact (100)	*S. aureus* 22837	Animal	France
phi22838-1	44.7	33	68	63604-108352	Intact (100)	*S. aureus* 22838	Animal	France
phi22841-1	44.7	33	66	63656-108404	Intact (100)	*S. aureus* 22841	Animal	France
phiSa52-2	73.2	33	74	40846-114049	Intact (140)	*S. aureus* Sa52	Animal	Denmark
phiRd.3-2	57.2	33.99	59	26855-84075	Intact (110)	*S. aureus* Rd.3	Animal	Germany
phiRd.9-1	70.2	33.2	69	4314-74563	Intact (140)	*S. aureus* Rd.9	Animal	Germany
phiRd.60-3	67.6	33.6	63	7053-74719	Intact (130)	*S. aureus* Rd.60	Animal	Germany
phi23237-1	44.7	33	68	63656-108404	Intact (100)	*S. aureus* 23237	Animal	France
phiC3489-2	29.2	35.38	38	2-29221	Intact (140)	*S. aureus* C3489	Animal	Spain
phiC5086-2	44.5	33.11	65	320203-364773	Intact (120)	*S. aureus* C5086	Animal	Spain

Table S11. Distribution of prophages of *S. aureus* into different clades obtained from the phylogenetic analysis.

Clade	Sub Clade	Phage Genomes	Common CDS features
Clade 1	-	phiG5-3, phi09-00736-1, phi08-01059-1, phiH14-1, phiH8-1, phiAB333-3, phiP333-3, phiH3-2, phiSA52-2, phiVET1913R-2, phiVET1914R-1, phiFP_N5208-2, phiFP_N5203-1, phiRD.9-1, and phiVET1915R-3	Phage hypothetical proteins (HP), phage proteins (PP), tail protein (TP), tail tape measure protein (TTmP), major tail protein (mTP), major tail protein, head-tail joining protein (HTJP), DNA packaging protein (DNA_PP), capsid protein (CapP), caseinolytic protease (Clp), portal protein (portP), terminase large subunit (TerL), terminase small subunit (TerS), HNH endonuclease, transcriptional activator rinB (TransA), virulence-associated E family protein (VirE), and DNA-binding protein.
Clade 2	-	phiHD1410-2, phiCM60-3, phiCM112-3, and phiCM124-1	Hypothetical proteins (HP), phage protein (PP), holin, tail fiber protein, N-acetylglucosaminidase (N-AGA), minor structural protein (mSP), and transcriptional activator rinB (TransA)
Clade 3	3A	phi22825-1, phi22838-1, phi22835-2, phi22841-1, phi22837-1, and phi23237-1	Hypothetical proteins (HP), phage proteins (PP), tail protein (TP), tail tape measure protein (TTmP), major tail protein (mTP), head-tail joining protein (HTJP), DNA packaging protein (DNA_PP), terminase small subunit (TerS), capsid protein (CapP), Clp protease, portal protein (portP), terminase large subunit (TerL), HNH endonuclease, transcriptional activator (TransA), PVL, antirepressor protein (AntiR), integrase, phospholipase C (Hlb), leukocidin/hemolysin toxin family protein (LukG/Hlg), minor structural proteins (mSP), holin, amidase, staphylokinase (SAK), SH3 domain-containing protein, Transposase binding protein, dUTPases, staphylococcal complement inhibitor (SCIN), chemotaxis inhibitory protein (CHIPS), transcriptional regulator (TransR), PemK like phage protein, and leukocidin/lemolysin toxin family protein (LukH).
	3B	phiI3-2, phiSA14-639-2, phiSA13-192-2, phiC5086-2, phiAB333-4, phiP333-4, phiG5-5, phiCM117-3, and phiH14-2	Hypothetical proteins (HP), phage proteins (PP), SCIN, SH domain-containing protein, chemotaxis inhibitory protein (CHIPs), amidase, holin, minor structural protein (mSP), tail protein (TP), tail tape measure protein (TTmP), major tail protein (mTP), head-tail joining protein (HTJP), DNA packaging protein (DNA_PP), terminase small subunit (TerS), capsid protein, Clp protease, portal protein, terminase large subunit (TerL), HNH endonuclease, transcriptional activator (TransA), dUTPase, PVL, transposase-associated ATP/GTP binding protein, and transcriptional regulator.
	3C	phiG4-3	-
	3D	phiSA13-192-1	-
Clade 4	-	phiI3-1, and phiHD1410-1	Hypothetical proteins, cyclase enzyme, metE, metH, bifunctional cystathionine gamma-lyase/gamma-synthase, parB, MscS family small conductance mechanosensitive

			ion channel protein, YchF, RpsF, RpsR, single-strand DNA binding protein (SSBP), PemK-like growth inhibitor, pathogenicity island protein (integrase), DNA-binding protein, DNA-binding protein, pathogenicity island DNA-binding protein, bovine pathogenicity island protein, mobile element-associated protein, primase, pathogenicity island protein, phage protein, mobile element-associated protein, spore coat protein, terminase small subunit (TerS), abortive infection bacteriophage resistance protein, integrase, YxeA family protein, and secreted protease inhibitor.
Clade 5	5A	phiH15-2, phiHD1410-3, phiH13-3, phiH1-3, phiH3-1, phiH4-3, and phiH2-3	Hypothetical proteins (HP), phage proteins (PP), transcriptional activator (TransA), HNH endonuclease, terminase large subunit (TerL), portal protein (portP), Clp protease, capsid protein (CapP), terminase small subunit (TerS), DNA packaging protein (DNA-PP), head-tail joining protein (HTJP), major tail protein (mTP), tail tape measure protein (TTmP), tail protein (TP), minor structural protein (mSP), amidase, staphylokinase (SAK), and SH3 domain-containing protein.
	5B	phiCM101-2, phi124-2, and phiH7-2	Phage proteins (PP), hypothetical proteins (HP), HNH endonuclease, terminase large subunit (TerL), head-tail joining protein (HTJP), tail tape measure protein (TTmP), tail protein (TP), minor structural proteins (mSP), holin, amidase, and staphylokinase (SAK).
Clade 6	-	phi53180-1-4, phi1111205429-1, phiRD.3-2, phiFP_N239-1, phi81629-1, phiH4-2, phiH1-1, phiH2-2, phiI3-4, and phiRD.60-3	Hypothetical proteins (HP), phage protein (PP), transcriptional activator RinB (TransA), virulence-associated E family protein (virE), VRR-NUC domain protein, HNH endonuclease, terminase small subunit (TerS), terminase large subunit (TerL), portal protein (portP), Clp protease, major capsid protein (CapP), DNA packaging protein (DNA_PP), head-tail joining protein (HTJP), major tail protein (mTP), tail protein (TP), and tail tape measure protein (TTmP).
Clade 7	7A	phiC3489-2, phiVET1912R-2, and phi2010-60-6511	Hypothetical proteins (HP), phage proteins (PP), transcriptional activator (TransA), terminase small subunit (TerS), terminase large subunit (TerL), Portal protein (portP), SPP1 family, Minor head protein (MHP), Head-tail adaptor (HTA), Tape measure proteins (TmP), Tail protein (TP), minor structural protein (mSP), N-acetylglucosaminidase (N-AGA), and tail fiber protein (TFP).
	7B	phiG5-2, phiCM117-1, phi909-4, and phi483-4	Hypothetical proteins (HP), phage proteins (PP), ssDNA-binding protein, PVL, dUTPase, transcriptional activator (TransA), terminase small subunit (TerS), Portal protein (portP), SPP1 family, head morphogenesis protein, terminase large subunit (TerL), tape measure protein (TmP), tail protein (TP), minor structural protein (mSP), N-acetylglucosaminidase (N-AGA), and tail fiber protein (TFP).

Table S12. Drug targets prioritization parameters and functional analysis of the protein targets

Protein Name	Protein ID	Gene	Length (aa)	Mo. Wt. (KDa)	Structural quality MHoLLine	Function
Biotin protein ligase	WP_000049913.1	*birA*	323	37.07	Very High	**MF**: Biotin-[acetyl-CoA-carboxylase] ligase activity. **BF**: Protein biotinylation
HPr kinase/phosphorylase	WP_000958224.1	*hprK*	310	34.48	Very High	**MF:** ATP binding Source, magnesium ion binding, phosphorelay sensor kinase activity, protein serine/threonine/tyrosine kinase activity, protein serine/threonine kinase activity. **BF:** Carbohydrate metabolic process, regulation of carbohydrate metabolic process
Thymidylate kinase	WP_001272126.1	*tmk*	210	24.06	Very High	**MF:** Thymidylate kinase activity. **BF:** Nucleotide biosynthesis
Phosphate acetyltransferase	WP_000774281.1	*pta*	328	34.98	Very High	**MF:** Phosphate acetyltransferase activity, **BF:** Acetyl-CoA biosynthetic process
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--L-lysine ligase	WP_000340119.1	*murE*	494	54.21	Very High	**MF:** UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase activity. **BF:** Cell cycle, Cell division, Cell shape, Cell wall biogenesis/degradation, Peptidoglycan synthesis
UTP--glucose-1-phosphate uridylyltransferase	WP_000721337.1	*gtaB*	288	32.49	Very High	**MF:** UTP:glucose-1-phosphate uridylyltransferase activity. **BF:** Enterobacterial common antigen biosynthetic process, UDP-glucose metabolic process
Fatty acid/phospholipid synthesis	WP_000239744.1	*plsX*	328	35.43	High	**MF:** Phosphate:acyl-[acyl carrier protein] acyltransferase activity. **BF:** Fatty

						acid biosynthetic process, phospholipid biosynthetic process
Pantothenate synthetase	WP_000163742.1	*panC*	283	31.44	Good	**MF:** ATP binding, pantoate-beta-alanine ligase activity. **BF:** Pantothenate biosynthesis

MF and **BF** represent Molecular function and biological function, respectively of the identified target proteins

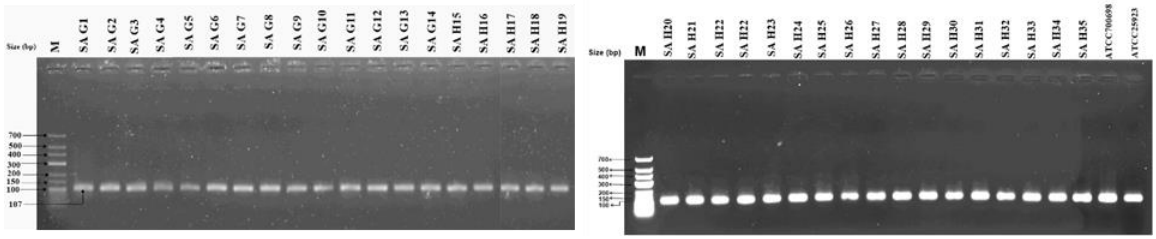


Fig. S1. Agarose gel electrophoresis of a species-specific sequence of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

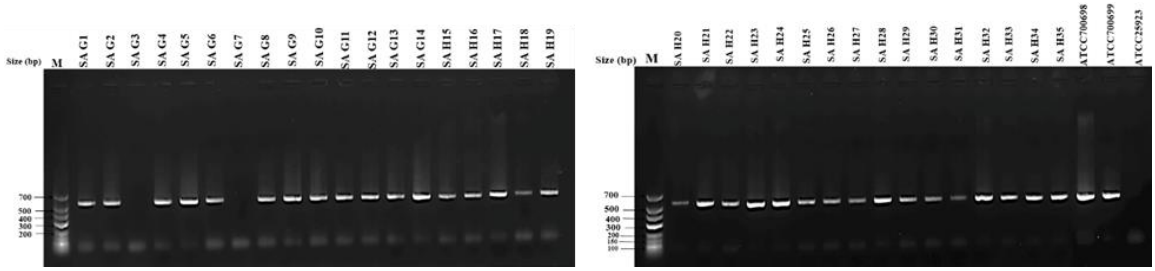


Fig. S2. Agarose gel electrophoresis of *mecA* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

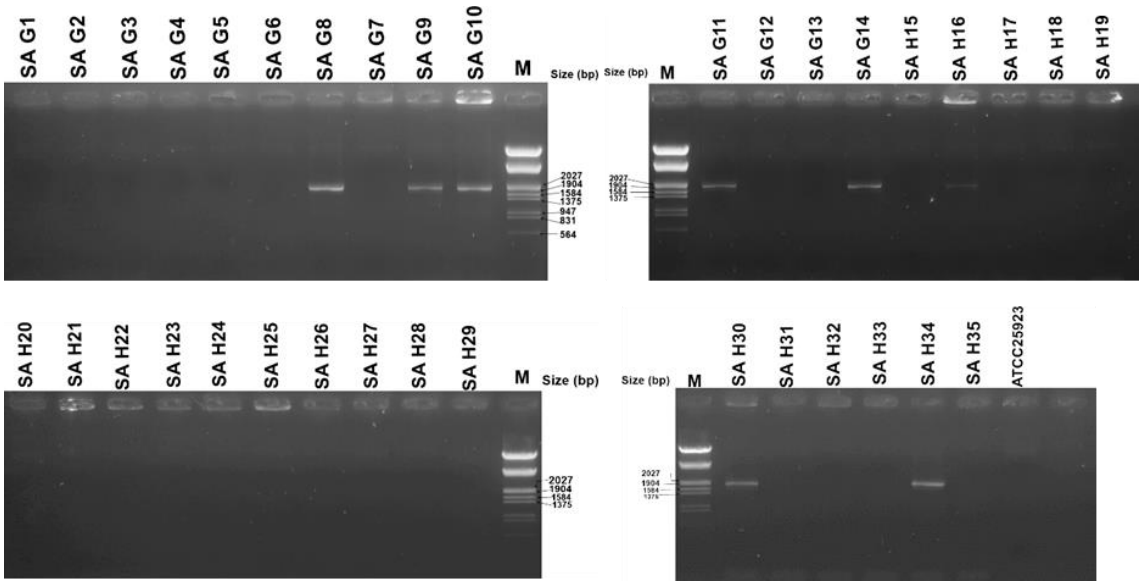


Fig. S3. Agarose gel electrophoresis of *pvl* gene of *S. aureus* clinical isolates. Lane M indicates the Lambda DNA digested with *EcoRI* and *HindIII* Marker 3 (Fermentas).

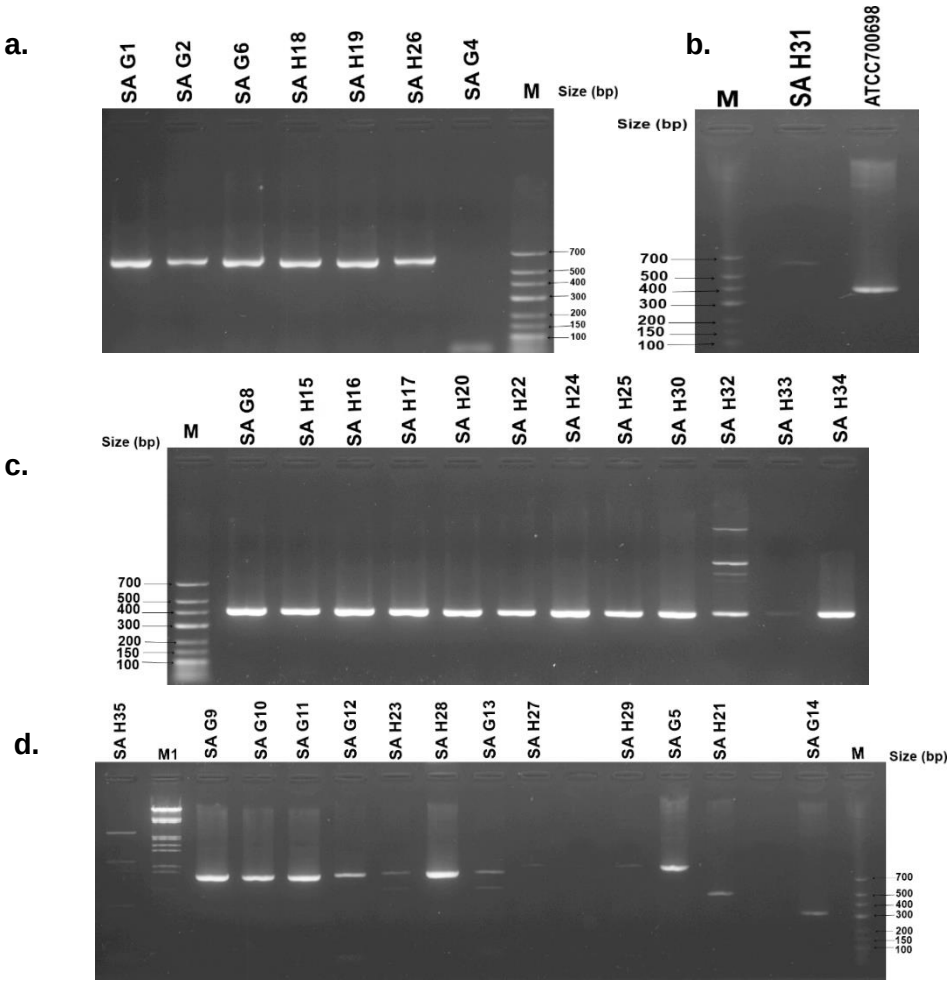


Fig. S4. Agarose gel electrophoresis of *SCCmec* gene of *S. aureus* clinical isolates. PCR amplified product of *SCCmec* type I **(a)** and **(b)**; *SCCmec* type II **(c)**; and *SCCmec* type IV and V **(d)**. Lane M indicates the low-range DNA ladder (Fermentas) and Lane M1 indicates the Lambda DNA digested with *EcoRI* and *HindIII* Marker 3 (Fermentas).

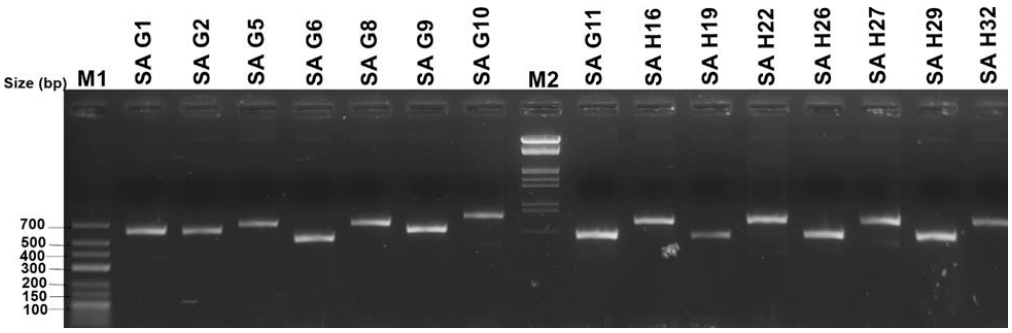


Fig. S5. Agarose gel electrophoresis of *coa* gene. Lane M1 indicates the low-range DNA ladder (Fermentas) and Lane M2 indicates the Lambda DNA digested with *EcoRI* and *HindIII* Marker 3 (Fermentas).

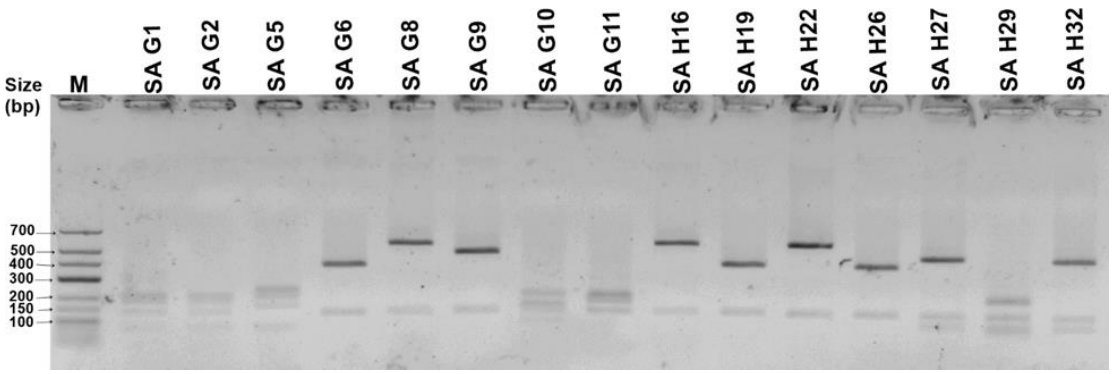


Fig. S6. Agarose gel electrophoresis of *coa*-*Hae*III-RFLP. Lane M indicates the low-range DNA ladder (Fermentas).

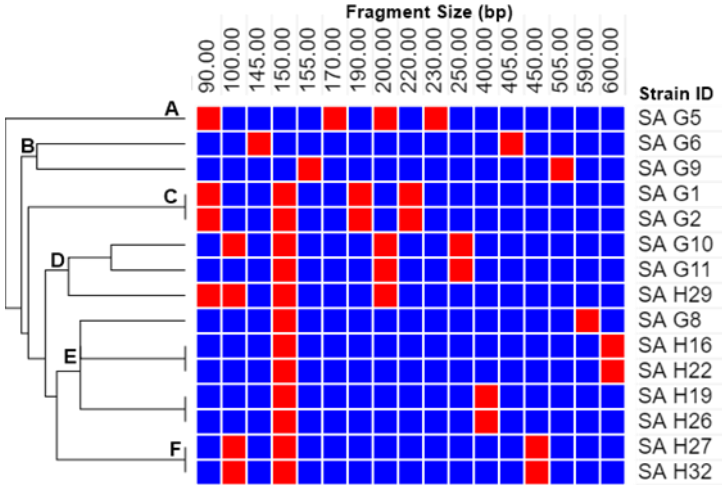


Fig. S7. Heat-map showing the similarity and difference of complex *coa*-*Hae*III RFLP banding pattern. The top labels indicate the fragment sizes (bp), Strain ID listed at right part of the panel, and clustering hierarchy demonstrated by the dendrogram at left.

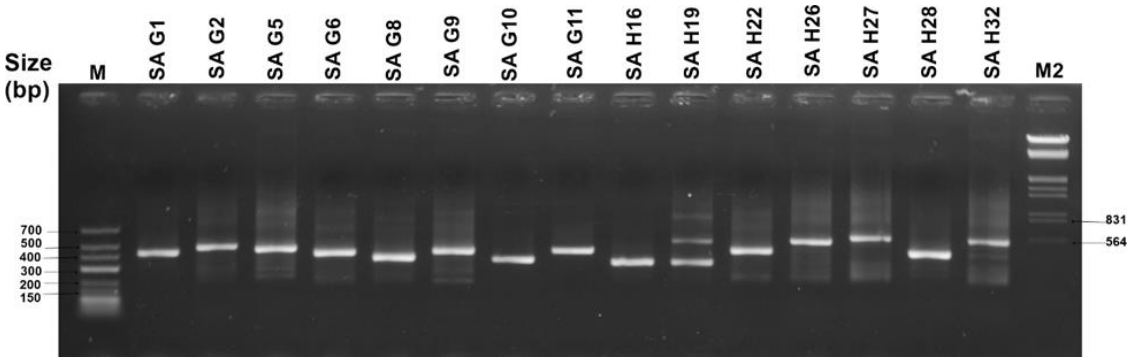


Fig. S8. Agarose gel electrophoresis of *spa* gene. Lane M indicates the low-range DNA ladder (Fermentas) and Lane M2 indicates the Lambda DNA digested with EcoRI and *Hind*III Marker 3 (Fermentas).

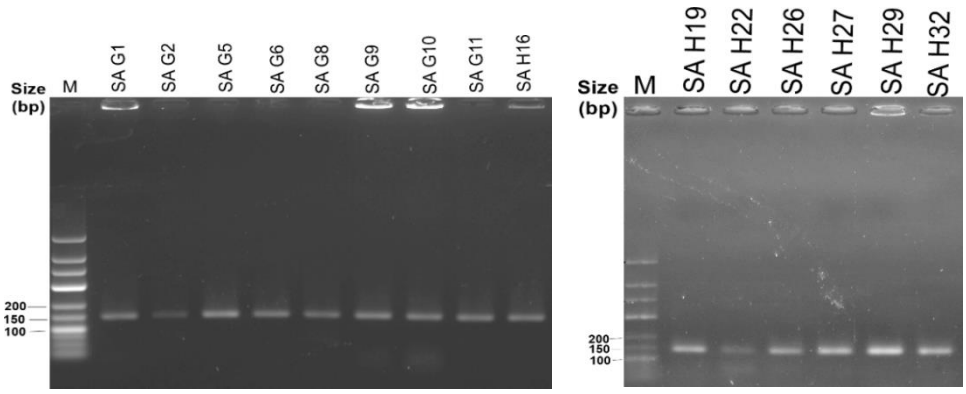


Fig. S9. Agarose gel electrophoresis of *icaA* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

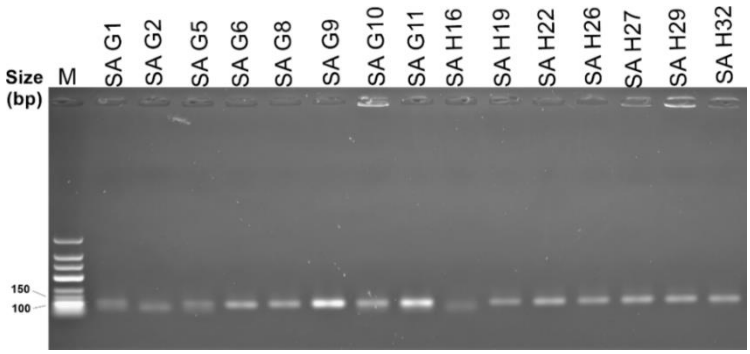


Fig. S10. Agarose gel electrophoresis of *icaB* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

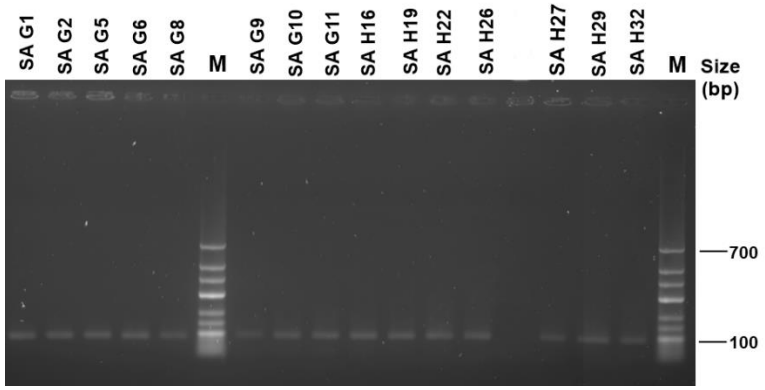


Fig. S11. Agarose gel electrophoresis of *icaC* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

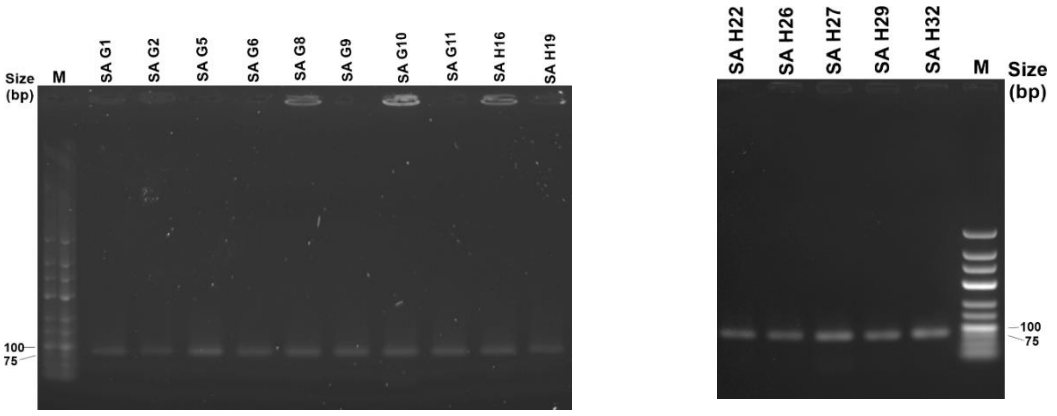


Fig. S12. Agarose gel electrophoresis of *icaD* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

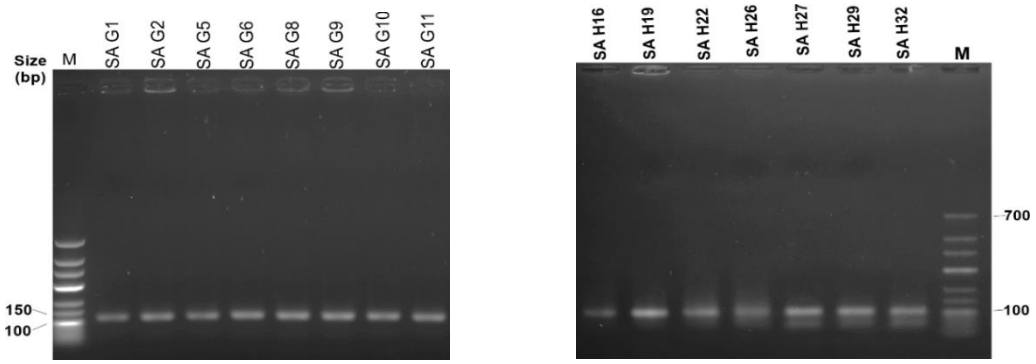


Fig. S13. Agarose gel electrophoresis of *icaR* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

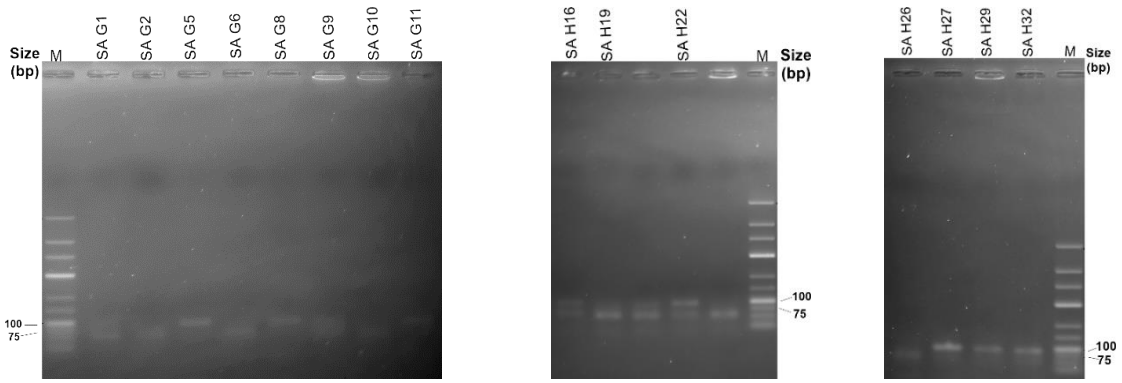


Fig. S14. Agarose gel electrophoresis of *fnaA* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

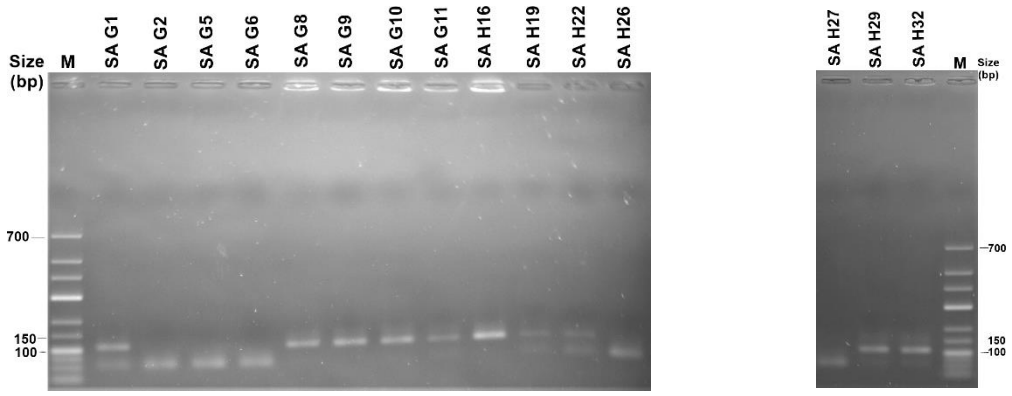


Fig. S15. Agarose gel electrophoresis of *fnaB* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

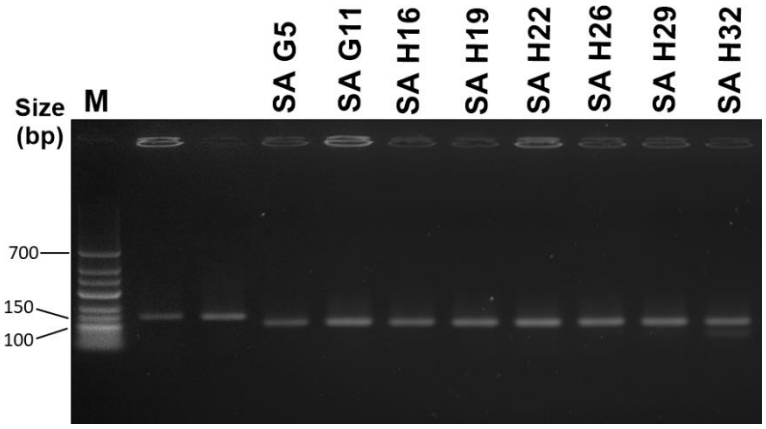


Fig. S16. Agarose gel electrophoresis of *cna* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

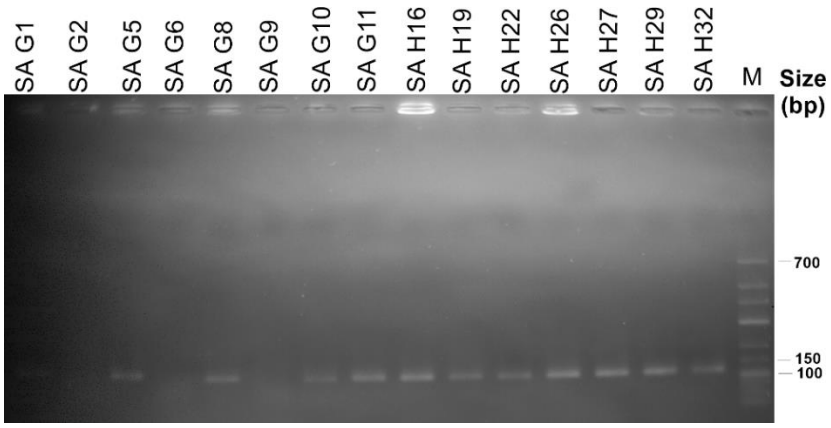


Fig. S17. Agarose gel electrophoresis of *clfA* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

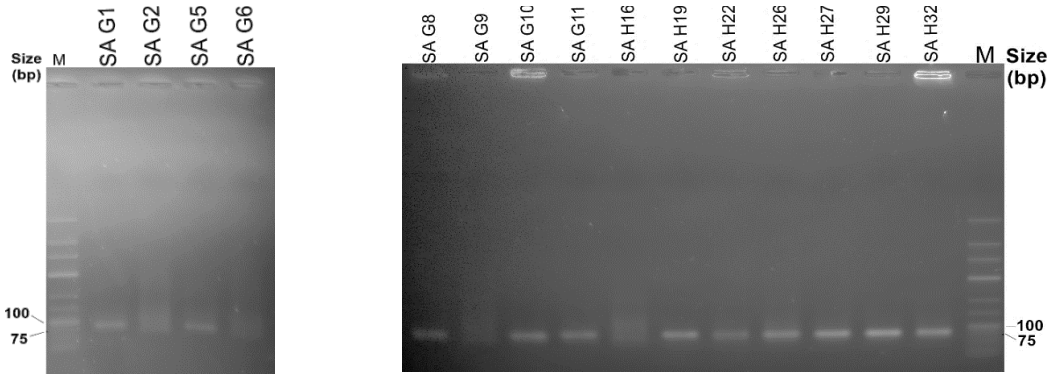


Fig. S18. Agarose gel electrophoresis of *clfB* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

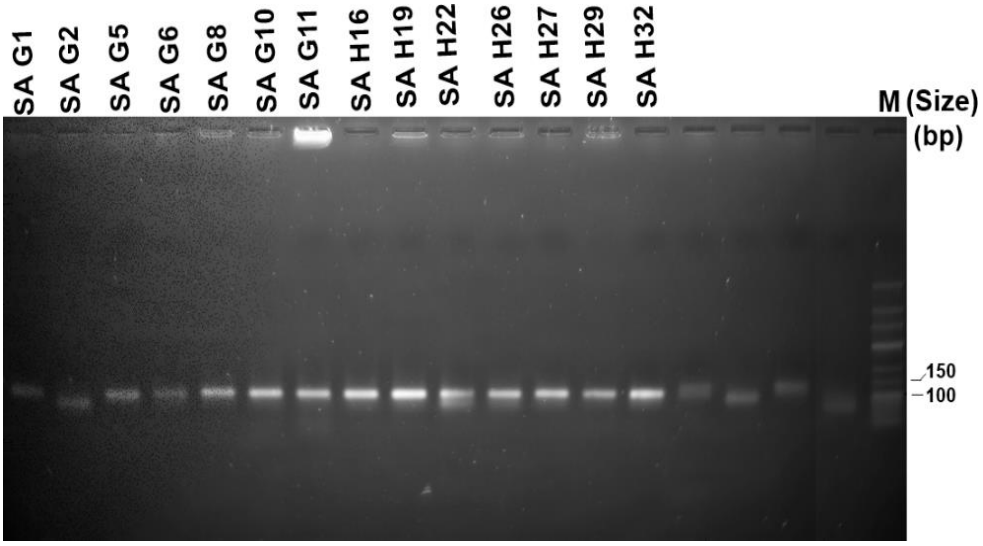


Fig. S19. Agarose gel electrophoresis of *ebps* gene of *S. aureus* clinical isolates. Lane M indicates the low-range DNA ladder (Fermentas).

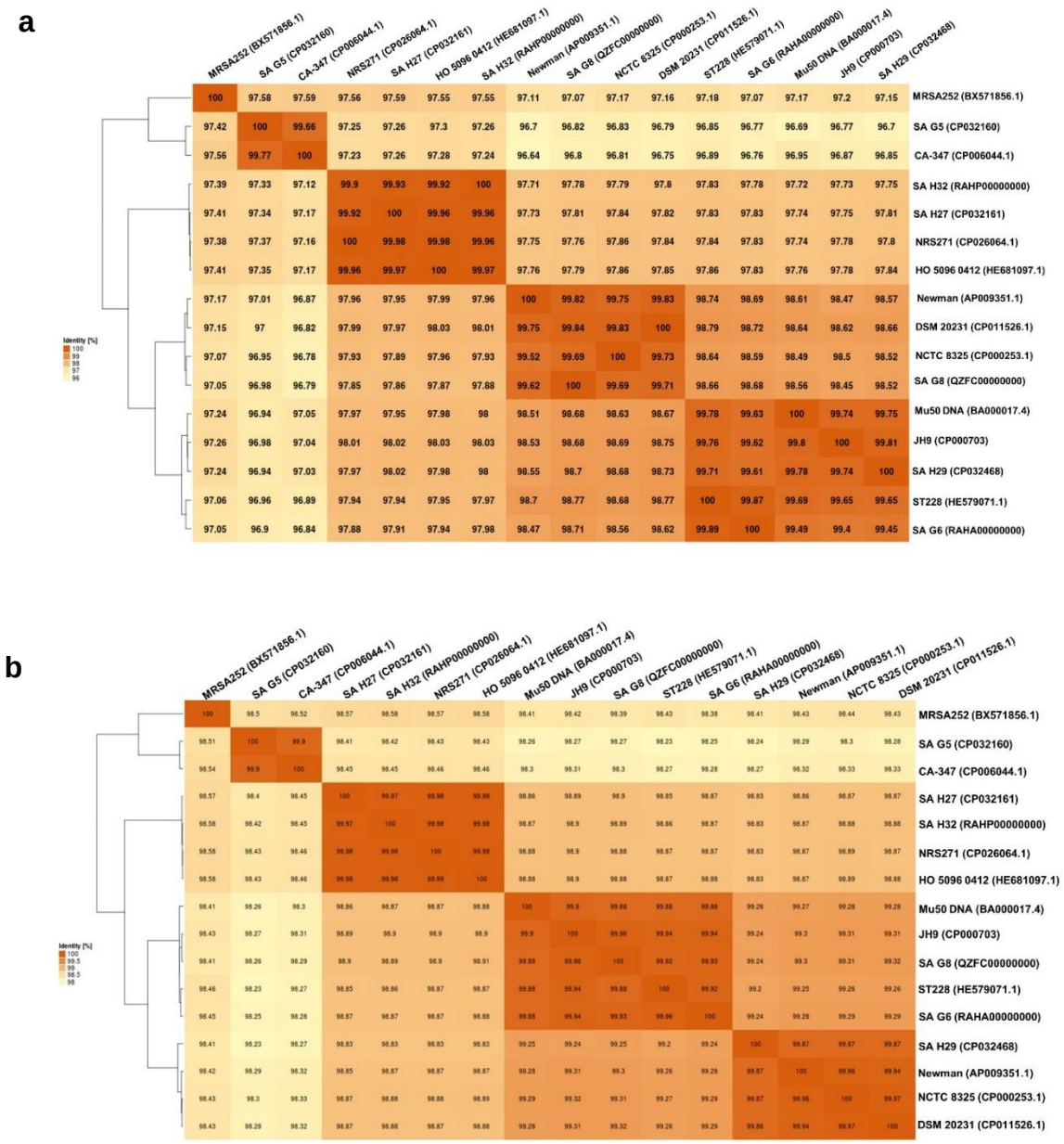


Fig. S20. Pairwise genome comparison of *S. aureus* and their closely related strains based on (a) ANI matrices, and (b) core genome ANI mean matrices.

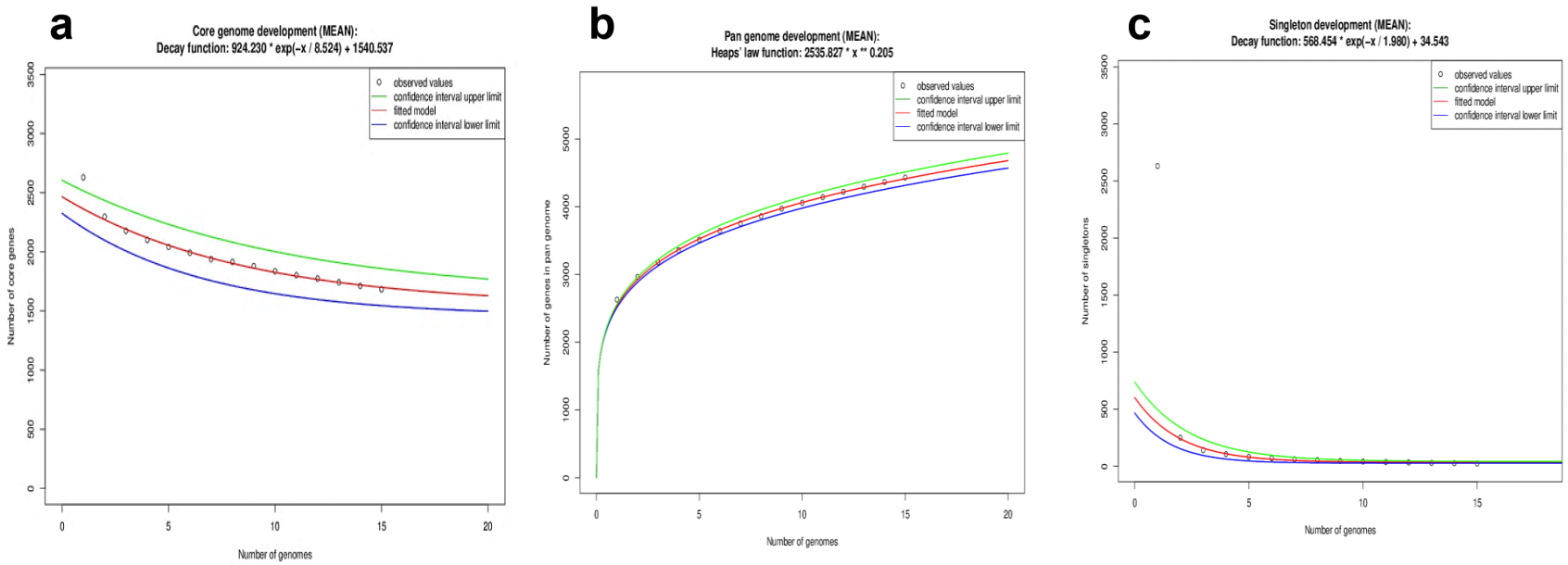


Fig. S21. Gene-set size statistical analysis of 15 *S. aureus* genomes. (a) The pan-genome development analysis suggested that increases with every additional *S. aureus* strain, indicating an open pan-genome (Heaps' law function: 2535.827* x **0.205). **(b)** Core-genome development of seven *S. aureus*. The calculated core genome will be around 1540.5 CDS, based on a decay function (924.230* exp(−x/8.524) +1540.537). **(c)** The singleton development analysis suggested that the pan-genome size will continue to expand at the rate of 34.5 genes per novel, representative genome based on a decay function (568.454* exp(-x/1.980+34.543).

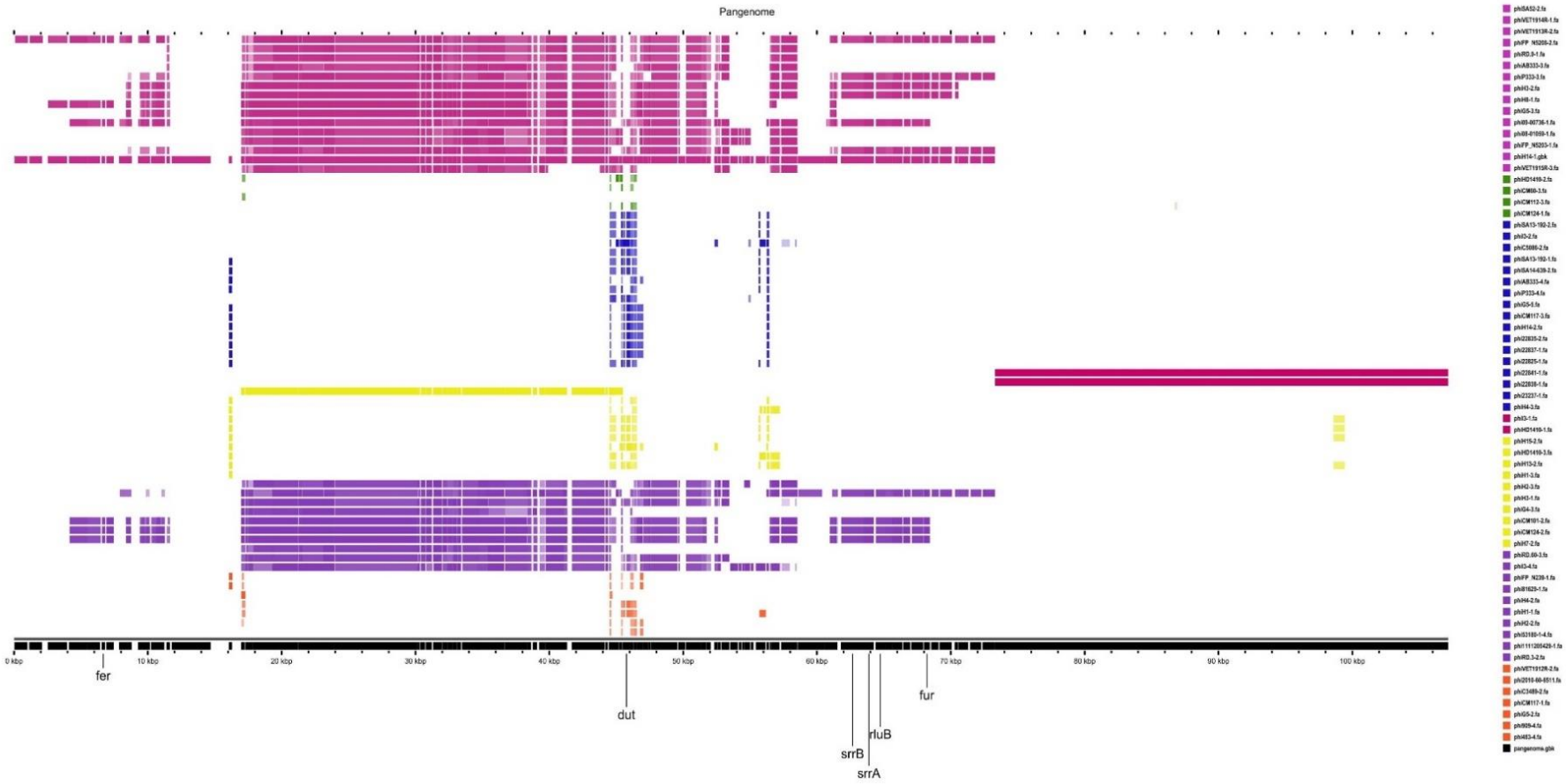


Fig. S22. Pan-genome analysis of prophage genomes of the *Siphoviridae* family. The pan-genome was generated in linear map with brown color below the background line (black), each linear genome sequence represents an individual putative prophage with color-coded based on the clade (clade 1 = marron, clade 2 = green, clade 3 = blue, clade 4=dark red, clade 5=yellow, clade 6=purple, and clade 7=red).

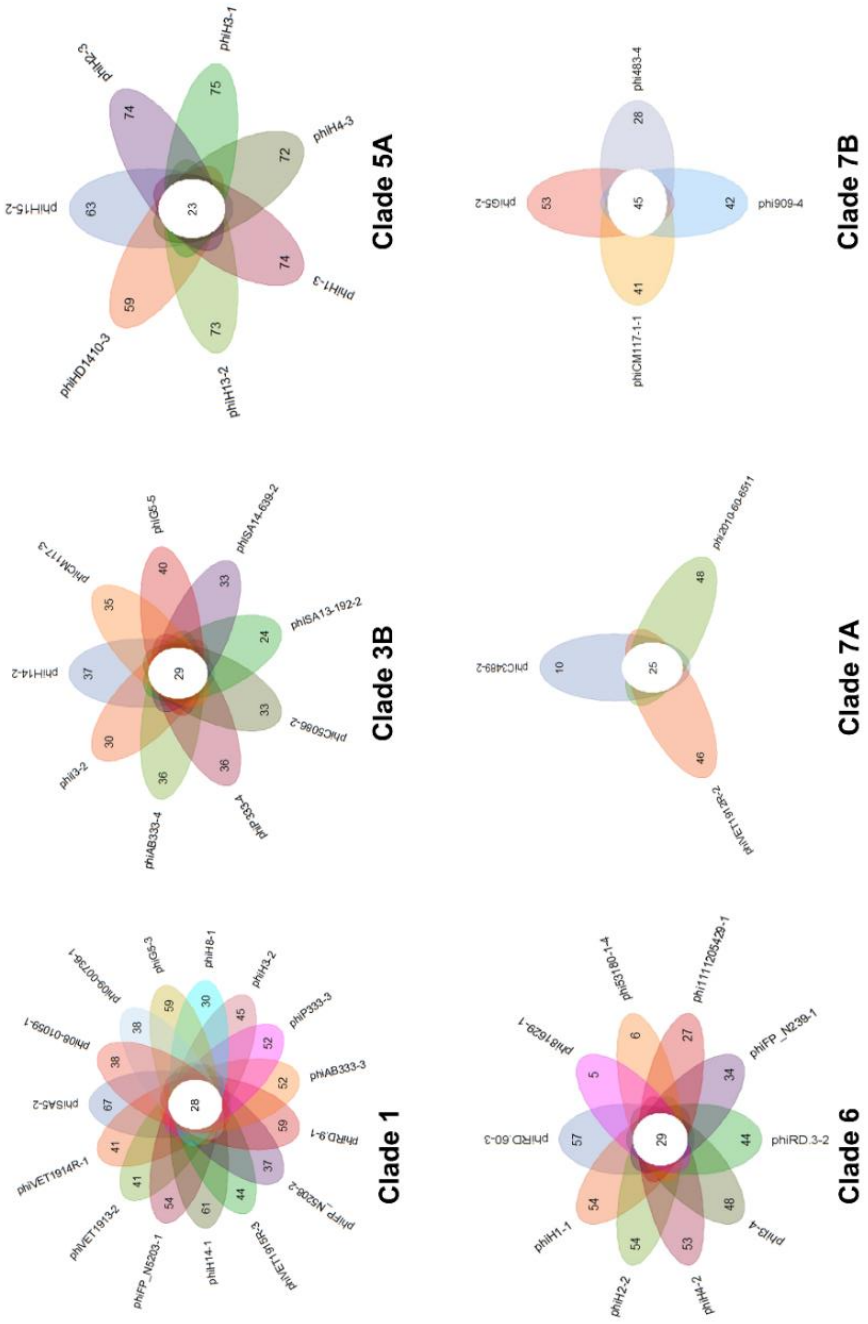


Fig. S23. Flower-plot showing the genes of core-genome (center) and accessory genes (petal). Each prophage is represented by an oval with a different color. In clade 1, the pan-genome has 168 CDS, and the core-genome has 28 CDS (Clade 1). In subclade 3B, the pan-genome has 111 CDS, and the core-genome has 29 CDS (Clade 3B). In subclade 5A, the pan-genome has 159 CDS, and the core-genome has 23 (Clade 5A). Clade 6 has a pan-genome comprised of 170 CDS, and the core-genome has 29 CDS (Clade 6). Subclade 7A formed pan-genome consists of 86 CDS, and core-genome size has 25 CDS (Clade 7A). In subclade 7B, the pan-genome has 115 CDS, and the core-genome consists of 45 (Clade 7B).

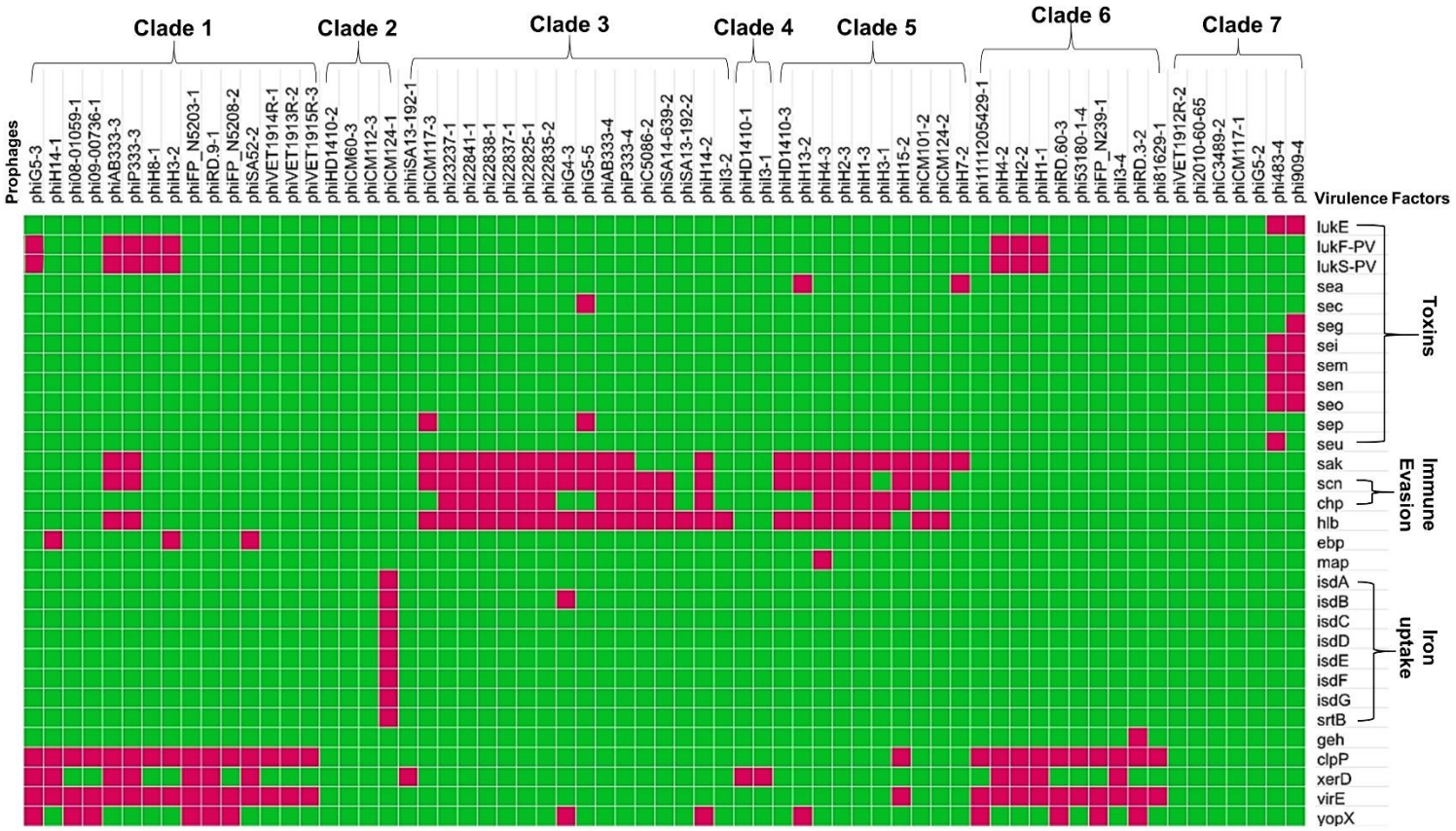


Fig. S24. Heatmaps showing the presence/absence of virulence genes in the prophages. The top labels indicate the prophage names and their clades (bold). The right labels indicate the virulence gene names and their categories (bold). The presence and absence of virulence genes are represented by red, and green colors, respectively.