Gu *et al.*
European Journal of Medical Research (2024) 29:474
https://doi.org/10.1186/s40001-024-02076-z

European Journal of Medical Research

RESEARCH

Open Access

Genome-wide comparative analysis of CC1 *Staphylococcus aureus* between colonization and infection

Feifei Gu[1†], Weiping He[2†], Dedong Zhu[1], Qian Zeng[1], Xinxin Li[3], Shuzhen Xiao[1], Yuxing Ni[1] and Lizhong Han[1*]

Abstract

Background *Staphylococcus aureus* is one of the most important bacteria in human colonization and infection. Clonal complex1 (CC1) is one of the largest and most important *S. aureus* CCs, and it is a predominant clone in *S. aureus* colonization and can cause a series of *S. aureus* infections including bloodstream infections. No studies on the relationship of CC1 *S. aureus* between colonization and infection have been published.

Methods To figure out if there are some significant factors in CC1 *S. aureus* help its colonization or infection, 15 CC1 *S. aureus* isolates including ten from colonization and five from bloodstream infections were enrolled in this study. Whole-genome sequencing and bioinformatics analysis were performed.

Results Virulence factor regulators XdrA, YSIRK signal peptide, CPBP family and OmpR family specifically found in infection isolates can promote virulence factors and enhance the pathogenicity of *S. aureus*. In addition, some significant differences in metabolism and human diseases were discovered between colonization and infection. Fst family of type I toxin–antitoxin system that mainly maintains stable inheritance was specifically found in CC1 *S. aureus* colonization isolates and might help *S. aureus* survive for colonization. No significant differences in genomic evolutionary relationship were found among CC1 *S. aureus* isolates between colonization and infection.

Conclusions Virulence factor regulators and metabolic state can promote CC1 *S. aureus* pathogenic process compared with colonization, and it seems that the strains of colonization origin cannot have pathogenic potential. Experimental confirmation and a bigger number of CC1 *S. aureus* strains are necessary for further study about the details and mechanism between colonization and infection.

Keywords *Staphylococcus aureus*, CC1, Colonization, Infection, Genome

Introduction

Staphylococcus aureus is an opportunistic pathogen that can colonize any part of the integument or mucous membranes, and nares is the most common site of colonization with approximately 30% of individuals permanently colonized [1]. Furthermore, *S. aureus* is a leading cause of bloodstream infections (BSIs), infective endocarditis as well as osteoarticular, skin and soft tissue, pleuropulmonary, and device-related infections [2]. *Staphylococcus aureus* causes the vast majority of skin and soft tissue infections (SSTIs), and it usually

†Feifei Gu and Weiping He have contributed equally to this study.

*Correspondence:
Lizhong Han
hanlizhong1107@163.com
[1] Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2] Department of Clinical Laboratory, Huaihe Hospital of Henan University, Kaifeng, Henan, China
[3] Department of Infectious Diseases, Research Laboratory of Clinical Virology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

progresses from asymptomatic colonization [3]. SSTIs can result in a range of severity from simple subcutaneous abscesses to severe necrotizing infections, and it may be both limb and life threatening as a potential source for bacteremia and metastatic abscesses [4]. The associations of colonization in different body sites with invasive infection have been proved in several studies [5]. It has been demonstrated that nasal *S. aureus* carriage might be a source of bacteremia; a substantial proportion of *S. aureus* bacteremia cases appear to be endogenous, since they originate from *S. aureus* colonized in nasal mucosa [6].

Nursing homes (NHs) for the elderly have been regarded as significant reservoirs for *S. aureus* and might play important roles in the spread of *S. aureus* in closed populations. It has been confirmed that the advanced aged were prone to intermittent *S. aureus* carriage, and the transmission of *S. aureus* among the participants living in a closed environment was revealed with a high genetic relatedness, suggesting its clonal spread in nursing homes [7]. A high prevalence of *S. aureus* colonization has been proved among nursing home residents in Shanghai in our previous studies, and clonal complex 1 (CC1) was the predominant clone among *S. aureus*-colonized residents both in 2014 (29.1%) and 2019 (36.3%) [8, 9].

CC1 was previously one of the five largest and most important *S. aureus* CCs including CC1, CC5, CC8, CC15 and CC97, and CC1 is known to contain various community-associated methicillin-resistant *S. aureus* (CA-MRSA) including the first PVL-positive strain MW2 [10]. Multilocus sequence type 1 (ST1), also known as pulsed-field type USA400, were among the most prominent CA-MRSA clone in the USA, and recently, ST1 has been sporadically reported in other regions, including China [11]. CC1 has been found as a dominant clone (86.0%) of MRSA in the fecal and nasal specimens from neonatal intensive care units (NICUs) patients in Japan [12], revealing a high prevalence of CC1 MRSA colonization and a threatening risk of MRSA infection remaining in NICU patients despite of appropriate control measures. CC1 has been recognized as a somewhat problematic clone in Europe as the clone European CC1-MRSA IV in recent years, and disseminated rapidly throughout Europe [13]. CC1 was still an infection clone until now found in a range of *S. aureus* infections even invasive infections and BSIs all over the world including in neonates in China [14–22]. In our latest study, CC1 was still found as an infection clone in MRSA burn wound infection in southeast China from 2013 to 2022 [23].

Considering the high prevalence of CC1 clone in *S. aureus* colonization and its importance in *S. aureus* infections, whether there is something special in CC1 clone might help its progress in colonization or infection is what we are concerned about. Unfortunately, there are no published studies on the CC1 *S. aureus* between colonization and infection. The transmission and spread of CC1 *S. aureus* might occur from colonization to infection with the high colonization prevalence of CC1 *S. aureus*. The aim of our study is to figure out if there is something significant in CC1 *S. aureus* between colonization and infection to aid their processes. The differences between CC1 *S. aureus* colonization and infection could help us understand which ones can help CC1 *S. aureus* colonize or enhance pathogenic potential and lead to infections. The factors specifically identified in colonization or infection may contribute to the development of control strategies regarding decolonization or infection prevention.

Materials and methods

Study isolates

Based on our previous studies on *S. aureus* colonization among NH residents in Shanghai, five CC1 *S. aureus* isolates in 2014 and five in 2019 refer to *S. aureus* colonization were enrolled in this study named G2014 (group 2014) and G2019 (group 2019), respectively. These ten CC1 *S. aureus* isolates in total were named GCOL (group colonization).

S. aureus isolates refer to infection were collected from *S. aureus* bloodstream infections from 2017 to 2019 in our previous studies. Five CC1 *S. aureus* isolates causing bloodstream infection were enrolled in this study named GBSI (group bloodstream infection). CC1 is not common in *S. aureus* bloodstream infection as other clones, and a limited number of five CC1 *S. aureus* infection isolates were all we found in *S. aureus* bloodstream infection in Shanghai during these years.

This study was approved by Ethics Committee of Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine. The strains enrolled in this study were collected from our previous studies [8, 9, 17] which all have been approved by Ethics Committee of Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine, and in this study we only focused on the bacteria without any more intervention. All methods in this study were carried out in accordance with relevant guidelines and regulations.

MLST and *spa* typing

Multilocus sequence typing (MLST) and *spa* typing were performed by polymerase chain reaction (PCR) on all *S. aureus* isolates collected according to the guidelines (https://pubmlst.org/, http://spa.ridom.de/index.shtml). CC1 and *spa* type t127 *S. aureus* isolates were identified, and 15 isolates were included for whole-genome sequencing in this study.

mecA detection and SCC*mec* typing

MRSA was screened by cefoxitin (30 μg) using the disk diffusion method following the Clinical and Laboratory Standards Institute guidelines [24]. *mecA* detection was performed on all *S. aureus* isolates to confirm the existence of MRSA. SCC*mec* types of MRSA were determined by the method as described previously [25].

Genomic DNA extraction

The CC1 *S. aureus* strains were streaked on Columbia blood agar and cultured at 37 ℃ overnight. A single colony on the plate was inoculated into 50 ml of LB liquid medium and cultured at 37 ℃ for approximately 16 h at 200 rpm. The cell biomass was harvested after 10 min centrifugation at 4000 rmp. Genomic DNA of the strains was extracted by Wizard® Genomic DNA Purification Kit (Promega), and purified genomic DNA was quantified by TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA). High quality DNA (OD260/280 = 1.8–2.0, > 1 μg) was used for further research.

Library construction and genome sequencing

The draft genome sequence analysis was carried out using the Hiseq X Ten sequencing platform (Major-Bio Co., Shanghai, China). Briefly, DNA samples were sheared into 400–500 bp fragments using a Covaris M220 Focused Acoustic Shearer according to the manufacture's protocol. Illumina sequencing libraries were prepared from the sheared fragments using the NEXTflex™ Rapid DNA-Seq Kit. 5′ primer ends were first end-repaired and phosphorylated. Next, the 3′ ends were A-tailed and ligated to sequencing adapters. The third step is to enrich the adapters-ligated products using PCR. The prepared libraries then were used for paired-end Illumina sequencing (2 × 150 bp) on an Illumina HiSeq X Ten sequencing platform.

Genome assembly and annotation

The data generated from Illumina platform were used for genomic analysis. The detailed procedures of genome assembly and annotation are as follows. Raw reads obtained after sequencing were filtered using a fastp software (version 0.19.6) [26] followed by assembly with SOPA de novo version 2.04. [27]. The final assembled genomes were submitted to GenBank with the accession number PRJNA847504. Glimmer [28] was used for CDS prediction, tRNA-scan-SE was used for tRNA prediction and Barrnap was used for rRNA prediction. The predicted CDSs were annotated from NR, Swiss-Prot, Pfam, GO, COG and KEGG database using sequence alignment tools such as BLASTP, Diamond and HMMER. Briefly, each set of query proteins were aligned with the databases, and annotations of best-matched subjects ($e < 10^{-5}$) were obtained for gene annotation.

Bioinformatics analysis

All of the bioinformatics and genomic analysis were performed using the online platform of Majorbio Cloud Platform (www.majorbio.com) from Shanghai Majorbio Bio-pharm Technology Co., Ltd. Phylogenetic tree was built using IQ-TREE (maximum-likelihood, ML) based on 27 housekeeping genes including *dnaG, nplT, infC, rpsJ, rplC, rplB, rpsS, rpsC, rplP, rplN, rplE, rplF, rpsE, rpsM, rpsK, rplM, nusA, frr, tsf, rpsB, rplS, pgk, pyrG, rplK, rplA, rplL, rpoB*. Model prediction software was jModelTest, the model selection criteria was BIC, and bootstraps was 1000. Average nucleotide identity (ANI) analysis was performed using the pyani software with ANIm method. Average aminoacid identity (AAI) analysis was performed using the CompareM software (*e*-value 10^{-3}, identity 30%, alignment length 70%). Pan-genome and core-genome analysis were performed using the PGAP software with GF method (*e*-value ≤ 10^{-5}, score ≥ 40, identity ≥ 50%, coverage ≥ 0, inflation 1.5).

Results

Isolates information

Fifteen CC1 (ST1, *spa* type t127) *S. aureus* isolates, including five from colonization in 2014 (G2014), five from colonization in 2019 (G2019), and five from bloodstream infection from 2017 to 2019 (GBSI), were enrolled in this study. Ten CC1 *S. aureus* colonization isolates (GCOL) were collected from nasal swabs of residents from seven NHs, and five CC1 *S. aureus* bloodstream infection isolates were collected from blood culture of patient with *S. aureus* bacteremia from three hospitals in Shanghai from 2017 to 2019. More details about the 15 CC1 *S. aureus* isolates are presented in Table 1.

Phylogenetic analysis

A phylogenetic tree of the 15 CC1 *S. aureus* isolates was built as Fig. 1 to infer and evaluate the evolutionary relationships between different genomes. Obviously, isolates did not group according to their origin (colonization/infection, resident/patient, or NH/hospital). However, isolates C11 and C12, which were collected from patients with *S. aureus* bacteremia in 2017 and 2018 in the same hospital, were grouped into the same sub-clade. ANI and AAI analysis were performed for genetic relationships at the genome-wide level, and no more differences have been found between different groups.

Table 1 Information of 15 CC1 *S. aureus* isolates from *S. aureus*-colonized residents in nursing homes and patients with *S. aureus* bloodstream infections in hospitals in Shanghai

Group	No. of isolates	Sample	Resident/patient	NH/hospital	Sampling date	CC (ST)	*spa* type	MRSA/MSSA	SCC*mec* type
GCOL									
G2014	C1	Nasal swab	Resident	NH1	2014-3-12	1	t127	MSSA	
	C2	Nasal swab	Resident	NH2	2014-3-25	1	t127	MSSA	
	C3	Nasal swab	Resident	NH3	2014-4-22	1	t127	MRSA	IV
	C4	Nasal swab	Resident	NH4	2014-4-23	1	t127	MRSA	IV
	C5	Nasal swab	Resident	NH5	2014-4-25	1	t127	MRSA	IV
G2019	C6	Nasal swab	Resident	NH6	2019-6-1	1	t127	MRSA	I
	C7	Nasal swab	Resident	NH6	2019-6-1	1	t127	MRSA	I
	C8	Nasal swab	Resident	NH7	2019-1-1	1	t127	MSSA	
	C9	Nasal swab	Resident	NH7	2019-1-1	1	t127	MSSA	
	C10	Nasal swab	Resident	NH7	2019-1-1	1	t127	MRSA	I
GBSI									
GBSI	C11	Blood	Patient	Hospital1	2017-5-20	1	t127	MSSA	
	C12	Blood	Patient	Hospital1	2018-11-13	1	t127	MSSA	
	C13	Blood	Patient	Hospital2	2019-6-22	1	t127	MRSA	I
	C14	Blood	Patient	Hospital3	2019-4-26	1	t127	MSSA	
	C15	Blood	Patient	Hospital1	2019-8-13	1	t127	MSSA	

Homologous genes analysis

Pan-genome and core-genome analysis have been performed to study the dynamic characteristics of bacterial genome and analyze the dynamic changes of bacterial genome in the process of evolution as shown in Fig. 2. The equation for estimating pan and core-genome size were given with the respective graphs, and the equations were represented in the figures, respectively. The pan-genome size was 3361 genes, and the core-genome size was 2156 genes.

The Venn diagrams were used to display the distribution of pan-genome, and the number of core genes and specific genes in different groups can be seen intuitively. There were 198 and 212 specific genes in GBSI and GCOL, respectively (Fig. 3a), and there were 76, 85 and 198 specific genes in G2014, G2019 and GBSI, respectively (Fig. 3b). After blasting and analyzing the specific genes in different groups, some possible significant proteins encoded by the specific genes might play important roles in colonization or infection as shown in Table 2. Fst family of type I toxin–antitoxin (TA) system which mainly probably helps maintaining stable inheritance was specifically found both in G2014 and 2019, but its detailed functions remain largely mysterious. It is worth noting that the vast majority of proteins uniquely discovered in GBSI are involved in the regulation of virulence factors, such as XdrA, YSIRK signal peptide, CPBP family and OmpR family. CRISPR–Cas (clustered regularly interspaced short palindromic repeat [CRISPR]-CRISPR-associated protein [Cas]) system, which was usually thought of as a genome editing tool, was specifically found in GBSI and might contribute to the pathogenesis of *S. aureus* human infections.

Genomic clustering analysis

Principal component analysis (PCA) and principal coordinates analysis (PCoA) were performed to study the close relationships between genomes of CC1 *S. aureus* isolates based on COG (Clusters of Orthologous Groups of proteins) and KEGG (Kyoto Encyclopedia of Genes and Genomes). As shown in Fig. 4, both in PCA and PCoA, the isolates in GBSI might not much closely compared with that in G2014 or/and G2019.

Correlation analysis was performed (Fig. 5) to evaluate the close relationships between CC1 *S. aureus* isolates based on COG and KEGG. The closer the correlation coefficient is to 1, the better the correlation between two isolates is. Isolate C14 appears to be much more different both in COG and in KEGG in correlation analysis compared with others. Besides, isolates C11, C12, C15 and C8 might be a little more different from other isolates in COG (Fig. 5a), and isolates C11, C15 and C8 might be more different from others in KEGG (Fig. 5b).

Functional analysis

According to the obtained abundance data of function genes and mobile elements, statistical comparison and hypothesis testing were carried out to evaluate the difference and significance level between different groups. The genes in different classifications of pan-genome

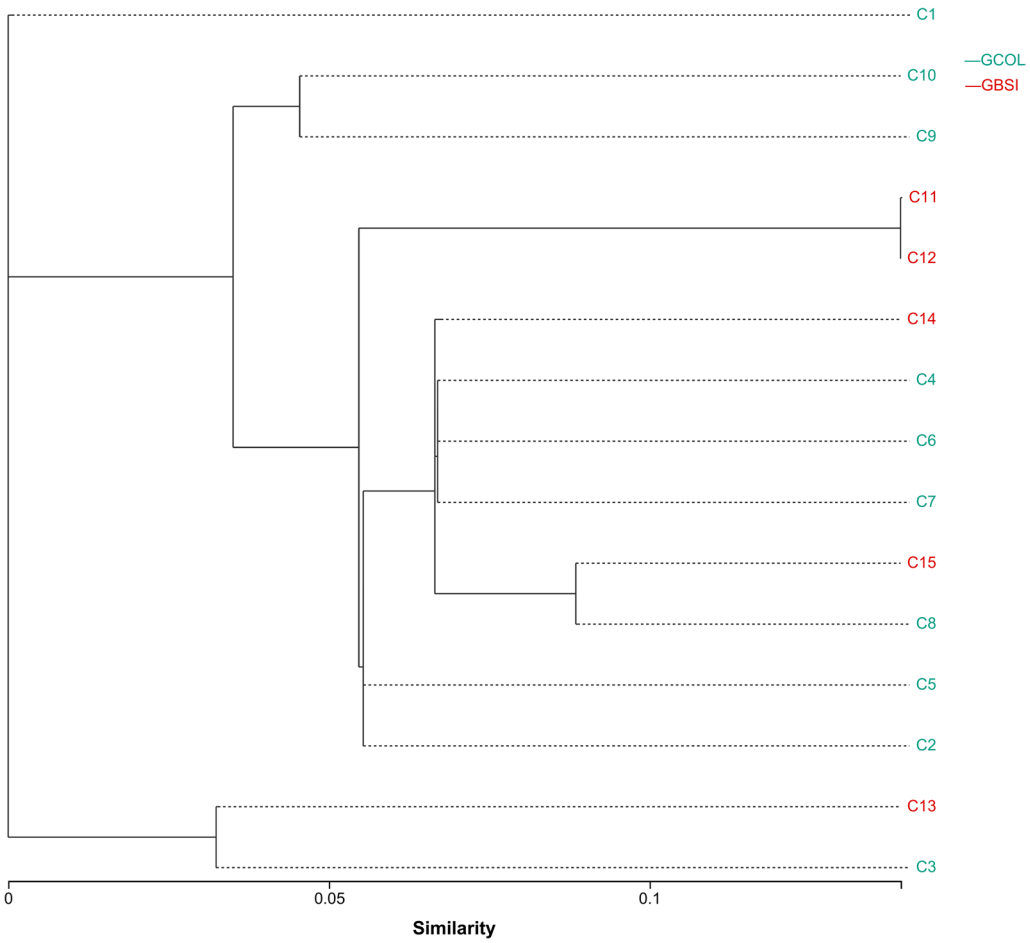

Fig. 1 The tree shows an evolutionary of 15 CCI *S. aureus* isolates based on 27 housekeeping genes, in which each branch represents one isolate, and the length of the branch is the evolutionary distance between two isolates

were annotated under different categories of COG and KEGG, and no significant difference was found including metabolism, information storage and processing, cellular processes and signaling, environmental information processing, genetic information processing, human diseases, organismal systems between GBSI and GCOL.

COG database comparison could be used in predicted proteins for functional annotation, classification, and protein evolution analysis. In COG comparative analysis, there might be significant different in metabolism ($P=0.0433$) between GBSI and GCOL, and no other differences were significantly found in COG.

The abundant pathway information in KEGG database can help us to understand the biological functions of genes at the system level, such as metabolic pathways, genetic information transmission, cytological processes, and some other complex biological processes. In KEGG comparative analysis, much more differences were significantly discovered in KEGG pathway level 1, level 2 and level 3 including in metabolism ($P=0.0319$) and human diseases ($P=0.0272$) between GCOL and GBSI. In addition, there were some significant differences in nucleotide metabolism ($P=0.0119$), pyrimidine metabolism ($P=0.0119$) and pentose phosphate pathway ($P=0.0361$) between G2014 and G2019 in KEGG pathway level 2 and level 3.

Discussion

S. aureus is one of the most infamous and widespread bacteria both in colonization and infections. On the one hand, about 15% to 32% of the population is thought to be nasally colonized by MSSA, and 1% to 3% is colonized by MRSA [29]. The rates of *S. aureus* colonization are typically higher in elder people as increase with advancing age [30], and the high prevalence of *S. aureus* colonization in the elderly in Shanghai has been proved as we studied before [8, 9]. Approximately 10–40% of people tested as outpatients or on admission are nasal carriages of *S. aureus* [6]. On the other hand, *S. aureus* can cause a hard-to-estimate number of uncomplicated skin

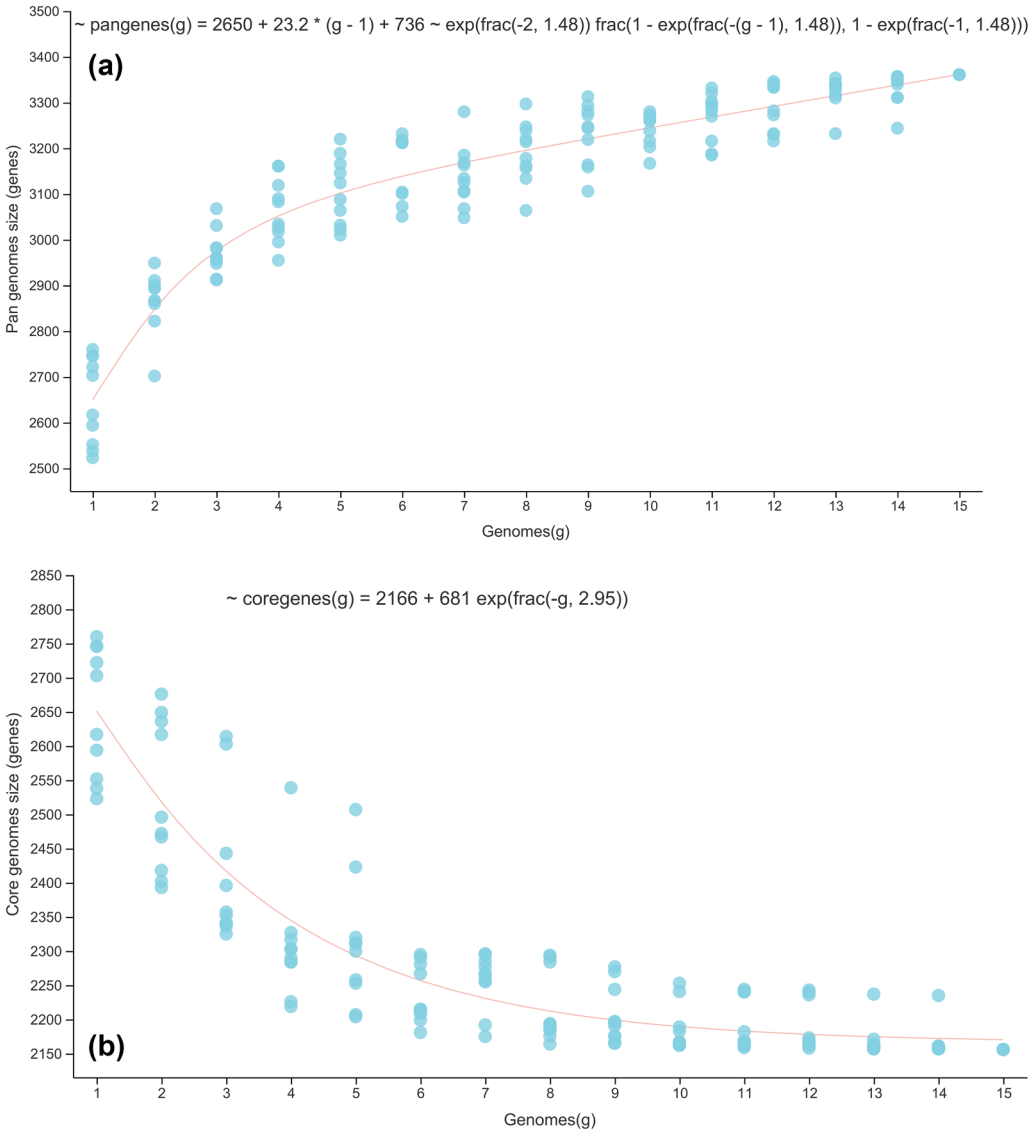

Fig. 2 Estimation of pan and core-genome size with the number of genomes. **a** Pan-genome; **b** core-genome. The abscissa is the number of genomes, the ordinate is the size of pan or core-genome. The equation for estimating pan and core-genome size according to PGAP is given with the respective graphs

infections and probably hundreds of thousands to millions of more severe, invasive infections including bloodstream infections globally per year [5]. The incidence of *S. aureus* bloodstream infections, leading high mortality rates (20–50%) despite proper treatment, has been rising in recent years in developed countries as reported [1].

CC1 is a predominant clone in *S. aureus* colonization, especially among the elderly people in nursing homes as we studied before. Meanwhile, CC1 is an infection clone in various types of *S. aureus* infections even in bloodstream infections and is one of the largest and most important *S. aureus* CCs. We believe that CC1 is an important clone deserved more attention considering the close relationship between colonization and infection that *S. aureus* infections usually arise from asymptomatic colonization [5].

According to the results of phylogenetic analysis, perhaps there is nothing remarkable about the evolutionary relationship between CC1 *S. aureus* colonization and infection. However, maybe it suggests that there is no obvious barrier in CC1 *S. aureus* between colonization and infection, and there are still continuous communication and transmission of CC1 *S. aureus* between colonization and infection, which means CC1 *S. aureus*

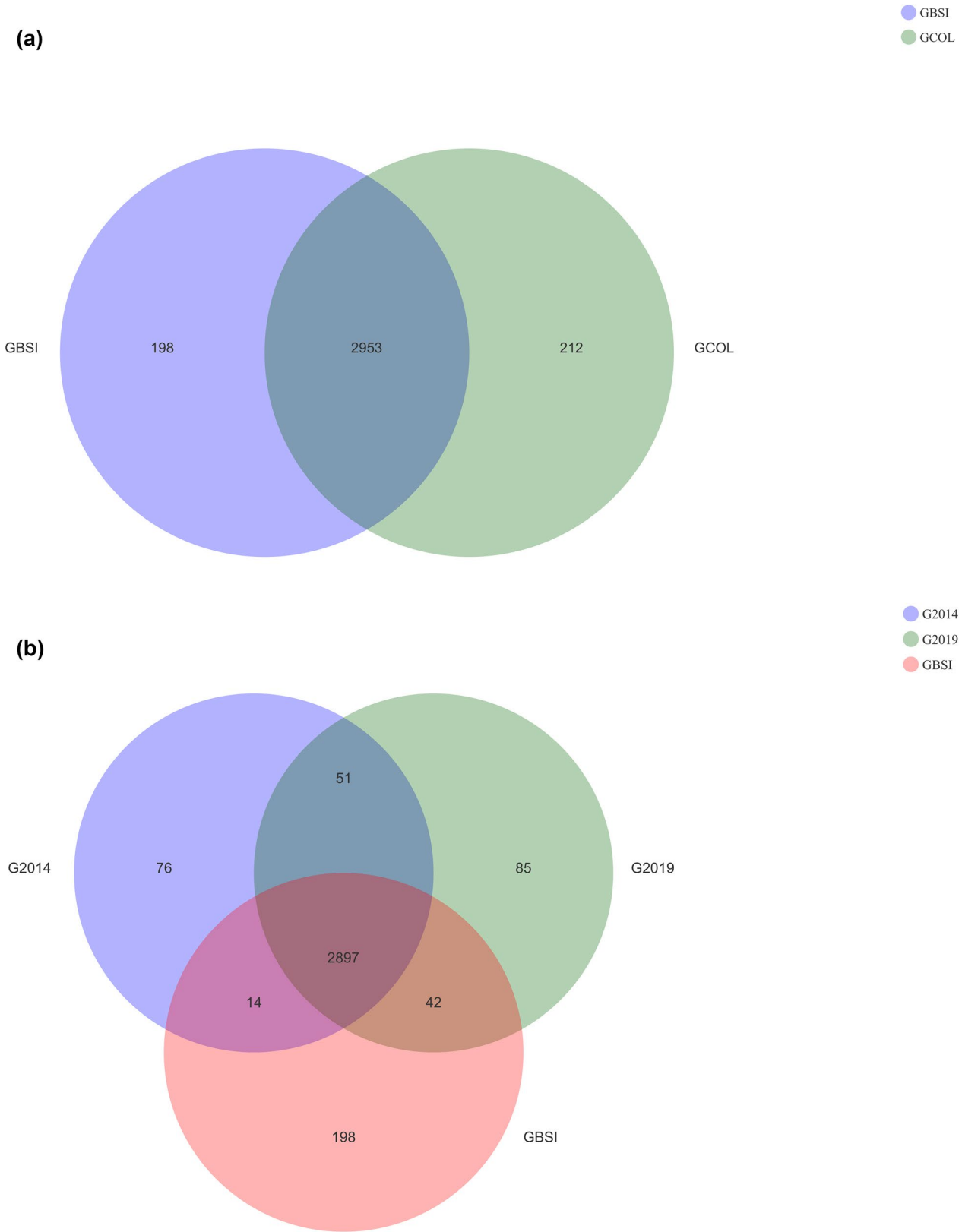

Fig. 3 The ellipse or petal of each color in the figure represents a group, the number of crossed petals indicates the number of shared homologous genes, and the number of individual petals indicates the number of homologous genes specific to the group. **a** GBSI and GCOL; **b** G2014, G2019 and GBSI

Table 2 Possible significant proteins encoded by the specific genes in different groups based on Venn analysis

Group	Protein	Function
GCOL (G2014 and G2019)	Fst family, type I toxin–antitoxin (TA) system	Plasmid-encoded TA systems including Fst systems: maintaining stable inheritance, remain largely mysterious; chromosomally encoded Fst systems: even more speculative [32]
GBSI	XRE-like DNA-binding regulator, A (XdrA)	A major activator of *spa* (encoding the virulence factor Staphylococcal protein A) expression that appears to act directly on the *spa* promoter [33]
GBSI	YSIRK signal peptide	Promotes Staphylococcal protein A (SPA) cross-wall targeting [36]
GBSI	CAAX proteases and bacteriocin-processing enzymes (CPBP) family	The proposed protease activity of bacterial CPBP members could be utilized in the maturation and secretion process of bacteriocins, and/or help conferring immunity against self-produced bacteriocins, details of their function remain elusive [37]
GBSI	OmpR family, DNA-binding response regulator transcription factor	Regulates the production of virulence factors by activating gene expression at the transcriptional level, such as adhesins, toxins, and immunomodulatory proteins; these virulence factors interact directly with host components to advance pathogenesis [38]
GBSI	Glucosyltransferase	Contribute to staphylococcal agglutination with fibrin fibrils in human plasma and the pathogenesis of sepsis; promote *S. aureus* replication in the bloodstream of mammalian hosts [45]
GBSI	Type III-A CRISPR–Cas system	Type III-A CRISPR–Cas systems can modulate bacterial genome stability by specifically preventing plasmid/phage transfer in a transcription-dependent manner, and may serve as a high-efficiency tool for deleting virulence or resistance genes in bacteria [42, 43]

colonization strains, may serve as endogenous reservoirs for overt clinical infections.

The homologous genes analysis, especially venn diagrams could provide more detailed information. After analyzing the specific genes and the proteins the genes encoded, a few proteins found might play important roles in *S. aureus* colonization or infection as presented in Table 2. Toxin–antitoxin (TA) system and other numerous genes in *S. aureus* could help detect and respond to diverse environmental stimuli such as nutrient starvation and stress to increase its fitness by altering their expression [31]. *S. aureus* genome contains 11 TA systems, and Fst family is one of type I TA system. Type I toxins have been demonstrated as bactericidal and can induce cell death in *S. aureus* [31]. However, the speculations on function of Fst family were still unclear. The role of plasmid-encoded TA systems in maintaining stable inheritance has been well established; nevertheless, TA system functions remain largely mysterious and the case of Fst system is no exception [32]. The function of chromosome-encoded Fst system is even more speculative. As is common for TA systems in general, the deletion of the Fst TA system has not been associated with any phenotypic effect on host cells [32]. Maybe the conditions have not yet been found under which the functions can be observed, or their functions are redundant with other bacterial genes. Even though no effect of Fst system has been found on host cells, its ability to stabilize hereditary material could help *S. aureus* survive stably in host for colonization.

Except glucosyltransferase and CRISPR–Cas system, the other proteins found in GBSI are regulators of virulence factors. XdrA (XRE-like DNA-binding regulator, A) shares similarity with XRE (xenobiotic response element) family helix-turn-helix, antitoxin-like proteins, and is the main activator which has been proved to bind directly to the *spa* promoter [33, 34]. XdrA regulates *spa* directly and does not join the interconnected regulatory network linking other well-characterized *spa* regulators [33]. Staphylococcal protein A (SPA) encoded by *spa* is a virulence factor whose function is to capture IgG molecules in the inverted orientation and therefore prevent phagocytosis of the bacterial cells by the host immune system [35]. Besides, YSIRK signal peptide, which is specifically found in GBSI as well, also has an effect on SPA that promotes SPA cross-wall targeting, though the mechanisms remain largely unknown [36].

The CPBP (CAAX proteases and bacteriocin-processing enzymes) family includes eukaryotic type II CAAX prenyl proteases and their related bacterial and archaeal homologs, and this family is widely represented in the current sequence database with more than 5,800 members present in all domains of life with most being bacterial [37]. Although most CPBP members are from bacteria, only a few have been experimentally studied, and the details of their function remain elusive. Based on available research, the function of many uncharacterized

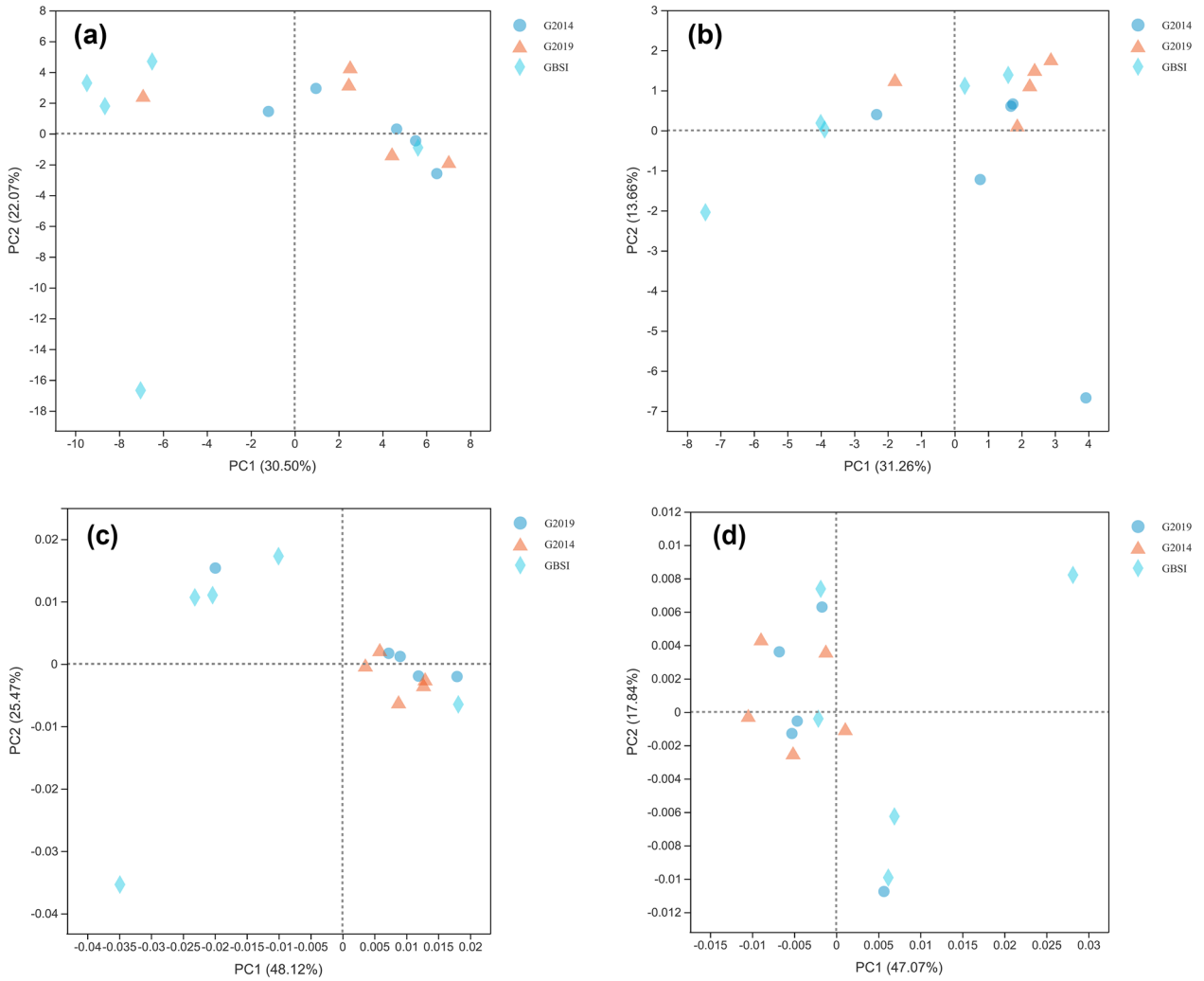

Fig. 4 The horizontal and vertical coordinates represent the two selected principal components, and the percentage represents the contribution value of the principal component to the difference in sample composition. The scale of horizontal and vertical axes is relative distance and has no practical significance. Points with different colors or shapes represent isolates of different groups. The closer the two isolate points are, the more similar the two isolates are. **a** PCA-COG; **b** PCA-KEGG; **c** PCoA-COG; **d** PCoA-KEGG

bacterial CPBP members might be related to bacteriocin maturation. It has been proposed that the protease activity of bacterial CPBP members could be utilized in the maturation and secretion process of bacteriocins, and/or help confer immunity against self-produced bacteriocins [37]. OmpR family, a DNA-binding response regulator transcription factor, regulates the production of virulence factors by activating gene expression at the transcriptional level, such as adhesins, toxins, and immunomodulatory proteins [38]. These virulence factors can cause host innate immune cells, such as macrophages and neutrophils, to release a variety of pro-inflammatory factors, leading to inflammatory damage, and induce cell apoptosis and necrosis of immune cells to escape the recognition and elimination of the innate immune system, leading to serious infectious diseases [39, 40].

CRISPR, known as an amazing natural immune system, is widespread in bacteria as well as in archaea and protects them against viral infections [41]. Type III-A CRISPR–Cas system was found in GBSI and the value of CRISPR–Cas system in adaptive immunity is described in detail in Table 2. Type III-A CRISPR–Cas systems can modulate bacterial genome stability by specifically preventing plasmid and/or phage transfer in a transcription-dependent manner, and may serve as a high-efficiency tool for deleting virulence or resistance genes in bacteria [42, 43]. The relevance of staphylococcal CRISPR element in pathogenesis warrants further investigation. A bigger

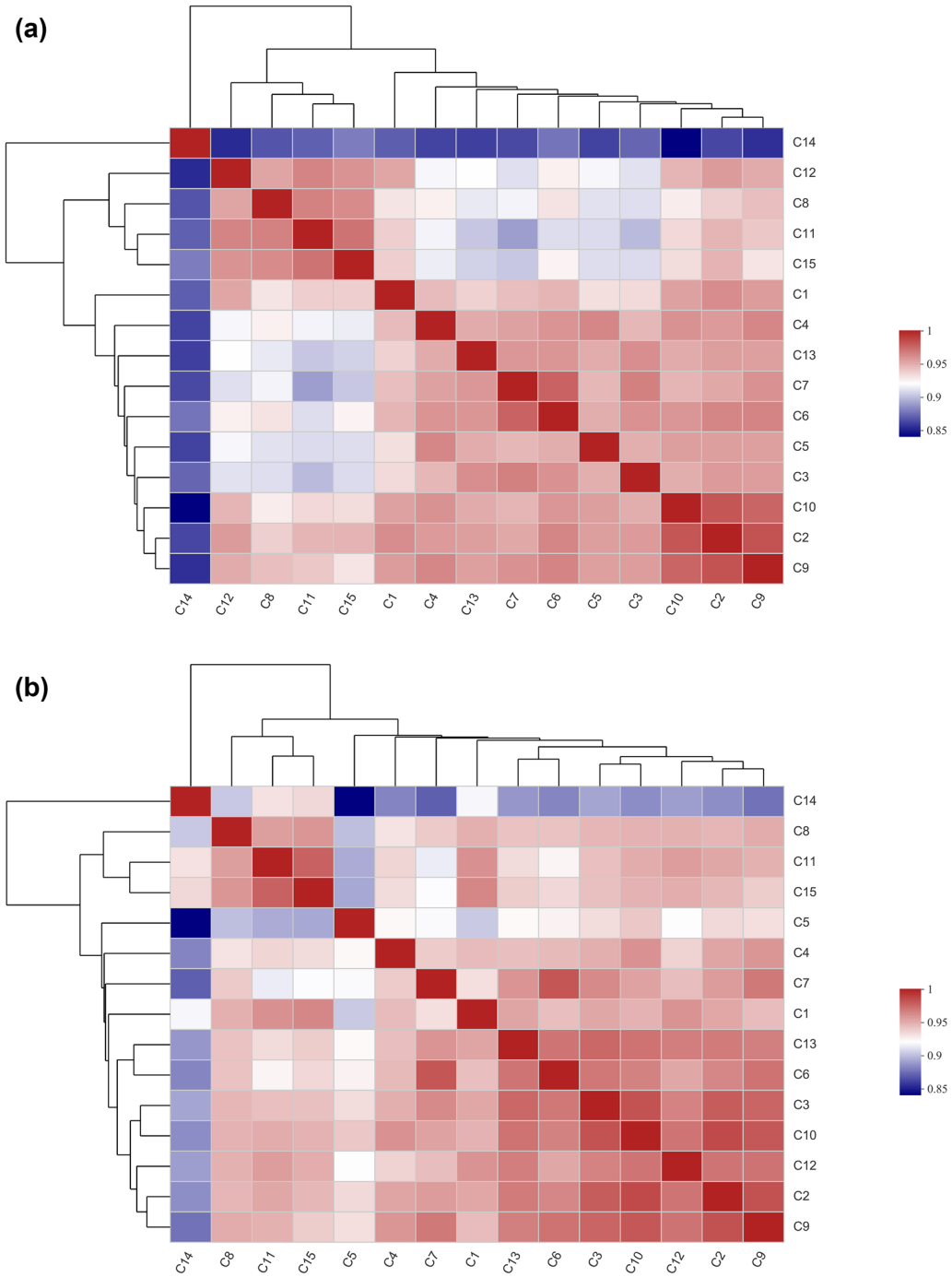

Fig. 5 The right and lower sides are isolate names, the left and upper sides are isolate clustering, and squares of different colors represent the correlation between the two isolates. The closer the color is to 1, the better the correlation between the two isolates is. **a** Correlation heat-map based on COG; **b** correlation heat-map based on KEGG

number of *S. aureus* strains from different sources especially those from infections origin are necessary for more research, hence additional features can be studied.

In functional analysis, some significant differences like metabolism and human diseases have been found in COG and KEGG between colonization and infection. The differences of metabolism were found both in COG and KEGG, suggesting maybe there are some metabolic differences indeed between colonization and infection. Manifestation of *S. aureus* disease is advanced by

expression of virulence factors, which is directly influenced by the metabolic state of the strain [44]. It makes sense that the differences in metabolism and virulence factor regulators we found between colonization and infection might have an impact on the virulence factors and make *S. aureus* more aggressive and leading to infections. The difference of human diseases in KEGG makes the results clearer that it seems the strains of colonization origin cannot have pathogenic potential. Microbial metabolomics and experimental studies should be carried out to learn more about information and details in metabolism between colonization and infection.

In conclusion, no significant differences in genomic evolutionary relationship have been found among CC1 *S. aureus* isolates between colonization and infection, suggesting there is still ongoing communication and transmission between colonization and infection. Fst family of maintaining stable inheritance might help CC1 *S. aureus* survive for colonization. Virulence factor regulators XdrA, YSIRK signal peptide, CPBP family and OmpR family, and the metabolic state of the strain, which can promote virulence factors and contribute to the pathogenesis of *S. aureus* infections, reveal maybe virulence factor regulators and metabolic state can help and accelerate *S. aureus* pathogenic process compared with colonization. According to the COG and KEGG analysis, it seems that the strains of colonization origin cannot have pathogenic potential. Experimental confirmation is extremely needed about the roles of Fst family, the virulence factor regulators, and the differences in metabolism between CC1 *S. aureus* colonization and infection.

Author contributions

Conceived and designed the study: F.G., W.H. and L.H. Performed the experiments: F.G., W.H. and D.Z. Analyzed the data F.G. and W.H. Contributed reagents/materials/analysis tools: Q.Z., X.L., S.X, D.Z and Y.N. Wrote the paper: F.G. Final approval of the submitted manuscript: All.

Funding

This study was supported by National Natural Science Foundation of China (Grant Numbers 82102452). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

All the sequence data in this study have been deposited in GenBank under BioProject ID PRJNA847504 (http://www.ncbi.nlm.nih.gov/bioproject/847504).

Declarations

Ethical approval and consent to participate

This study was approved by Ethics Committee of Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine, and the Review Committee removed the need for informed consent for this retrospective study which focused on bacteria and did not involve patient interventions.

Competing interests

The authors declare no competing interests.

Received: 30 August 2023 Accepted: 24 September 2024
Published online: 30 September 2024

References

1. Kwiecinski JM, Horswill AR. *Staphylococcus aureus* bloodstream infections: pathogenesis and regulatory mechanisms. Curr Opin Microbiol. 2020;53:51–60. https://doi.org/10.1016/j.mib.2020.02.005.
2. Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG Jr. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev. 2015;28:603–61. https://doi.org/10.1128/CMR.00134-14.
3. Dryden MS. Skin and soft tissue infection: microbiology and epidemiology. Int J Antimicrob Agents. 2009;34(Suppl 1):S2-7. https://doi.org/10.1016/s0924-8579(09)70541-2.
4. Cardona AF, Wilson SE. Skin and soft-tissue infections: a critical review and the role of telavancin in their treatment. Clin Infect Dis. 2015;61(Suppl 2):S69-78. https://doi.org/10.1093/cid/civ528.
5. Cheung GYC, Bae JS, Otto M. Pathogenicity and virulence of *Staphylococcus aureus*. Virulence. 2021;12:547–69. https://doi.org/10.1080/21505594.2021.1878688.
6. von Eiff C, Becker K, Machka K, Stammer H, Peters G. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study Group. N Engl J Med. 2001;344:11–6. https://doi.org/10.1056/nejm200101043440102.
7. Kasela M, et al. Transmission and long-term colonization patterns of *Staphylococcus aureus* in a nursing home. Int J Environ Res Public Health. 2020. https://doi.org/10.3390/ijerph17218073.
8. Zhang J, et al. Prevalence and molecular epidemiology of *Staphylococcus aureus* among residents of seven nursing homes in Shanghai. PLoS ONE. 2015;10: e0137593. https://doi.org/10.1371/journal.pone.0137593.
9. He WP, et al. Molecular characteristics and risk factor analysis of *Staphylococcus aureus* colonization put insight into CC1 colonization in three nursing homes in Shanghai. PLoS ONE. 2021;16: e0253858. https://doi.org/10.1371/journal.pone.0253858.
10. Dabul AN, Camargo IL. Clonal complexes of *Staphylococcus aureus*: all mixed and together. FEMS Microbiol Lett. 2014;351:7–8. https://doi.org/10.1111/1574-6968.12358.
11. Chen Y, et al. The novel fosfomycin resistance gene fosY is present on a genomic island in CC1 methicillin-resistant *Staphylococcus aureus*. Emerg Microb Infect. 2022;11:1166–73. https://doi.org/10.1080/22221751.2022.2058421.
12. Tsujiwaki A, et al. Epidemiology of methicillin-resistant *Staphylococcus aureus* in a Japanese neonatal intensive care unit. Pediatrics Int. 2020;62:911–9. https://doi.org/10.1111/ped.14241.
13. Earls MR, et al. Exploring the evolution and epidemiology of European CC1-MRSA-IV: tracking a multidrug-resistant community-associated methicillin-resistant *Staphylococcus aureus* clone. Microb Genom. 2021. https://doi.org/10.1099/mgen.0.000601.
14. Senok A, et al. Genotyping of methicillin resistant *Staphylococcus aureus* from the United Arab Emirates. Sci Rep. 2020;10:18551. https://doi.org/10.1038/s41598-020-75565-w.
15. Noone JC, et al. Molecular characteristics of *Staphylococcus aureus* associated prosthetic joint infections after hip fractures treated with hemiarthroplasty: a retrospective genome-wide association study. Sci Rep. 2020;10:16553. https://doi.org/10.1038/s41598-020-73736-3.
16. Monecke S, et al. Genotyping of methicillin-resistant *Staphylococcus aureus* from sepsis patients in Pakistan and detection of antibodies against staphylococcal virulence factors. Eur J Clin Microbiol Infect Dis. 2020;39:85–92. https://doi.org/10.1007/s10096-019-03695-9.
17. Gu F, et al. Antimicrobial resistance and molecular epidemiology of *Staphylococcus aureus* causing bloodstream infections at Ruijin Hospital in Shanghai from 2013 to 2018. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-63248-5.
18. Firoozeh F, Omidi M, Saffari M, Sedaghat H, Zibaei M. Molecular analysis of methicillin-resistant *Staphylococcus aureus* isolates from four teaching hospitals in Iran: the emergence of novel MRSA clones. Antimicrob Resist Infect Control. 2020. https://doi.org/10.1186/s13756-020-00777-8.

19. Dong Q, et al. Phenotypic and molecular characteristics of community-associated *Staphylococcus aureus* infection in neonates. Infect Drug Resist. 2020;13:4589–600. https://doi.org/10.2147/IDR.S284781.
20. Boswihi SS, Udo EE, AlFouzan W. Antibiotic resistance and typing of the methicillin-resistant *Staphylococcus aureus* clones in Kuwait hospitals, 2016–2017. BMC Microbiol. 2020. https://doi.org/10.1186/s12866-020-02009-w.
21. Vidal C, et al. Fulminant arterial vasculitis as an unusual complication of disseminated staphylococcal disease due to the emerging CC1 methicillin-susceptible *Staphylococcus aureus* clone: a case report. BMC Infect Dis. 2019. https://doi.org/10.1186/s12879-019-3933-3.
22. Monteiro AS, et al. Phylogenetic and molecular profile of *Staphylococcus aureus* isolated from bloodstream infections in Northeast Brazil. Microorganisms. 2019. https://doi.org/10.3390/microorganisms7070210.
23. Gu F, et al. A 10-year retrospective study of methicillin-resistant *Staphylococcus aureus* from burn wound infection in southeast China from 2013 to 2022. Front Microbiol. 2023. https://doi.org/10.3389/fmicb.2023.1301744.
24. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing. CLSI document M100-S29, Wayne, PA, USA 39. 2019.
25. Chongtrakool P, et al. Staphylococcal cassette chromosome mec (SCCmec) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCCmec elements. Antimicrob Agents Chemother. 2006;50:1001–12. https://doi.org/10.1128/aac.50.3.1001-1012.2006.
26. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
27. Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1:18. https://doi.org/10.1186/2047-217x-1-18.
28. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007;23:673–9. https://doi.org/10.1093/bioinformatics/btm009.
29. Sharara SL, Maragakis LL, Cosgrove SE. Decolonization of *Staphylococcus aureus*. Infect Dis Clin North Am. 2021;35:107–33. https://doi.org/10.1016/j.idc.2020.10.010.
30. Baldwin NS, et al. Prevalence of methicillin-resistant *Staphylococcus aureus* colonization in residents and staff in nursing homes in Northern Ireland. J Am Geriatr Soc. 2009;57:620–6. https://doi.org/10.1111/j.1532-5415.2009.02181.x.
31. Habib G, Zhu Q, Sun B. Bioinformatics and functional assessment of toxin-antitoxin systems in *Staphylococcus aureus*. Toxins. 2018. https://doi.org/10.3390/toxins10110473.
32. Weaver K. The Fst/Ldr family of type I TA system toxins: potential roles in stress response, metabolism and pathogenesis. Toxins. 2020. https://doi.org/10.3390/toxins12080474.
33. McCallum N, Hinds J, Ender M, Berger-Bachi B, Stutzmann Meier P. Transcriptional profiling of XdrA, a new regulator of spa transcription in *Staphylococcus aureus*. J Bacteriol. 2010;192:5151–64. https://doi.org/10.1128/JB.00491-10.
34. Oscarsson J. Regulatory role of proteins binding to the (protein A) and (Staphylococcal accessory regulator) promoter regions in NTCC 8325–4. Int J Med Microbiol. 2005;295:253–66. https://doi.org/10.1016/j.ijmm.2005.05.003.
35. Votintseva AA, et al. Prevalence of *Staphylococcus aureus* protein A (spa) mutants in the community and hospitals in Oxfordshire. BMC Microbiol. 2014;14:63. https://doi.org/10.1186/1471-2180-14-63.
36. Zhang R, et al. Spatial regulation of protein A in *Staphylococcus aureus*. Mol Microbiol. 2021;116:589–605. https://doi.org/10.1111/mmi.14734.
37. Pei J, Mitchell DA, Dixon JE, Grishin NV. Expansion of type II CAAX proteases reveals evolutionary origin of gamma-secretase subunit APH-1. J Mol Biol. 2011;410:18–26. https://doi.org/10.1016/j.jmb.2011.04.066.
38. Fan R, et al. The effects of L-arginine on protein stability and DNA binding ability of SaeR, a transcription factor in *Staphylococcus aureus*. Protein Expr Purif. 2021;177:105765. https://doi.org/10.1016/j.pep.2020.105765.
39. Zecconi A, Scali F. *Staphylococcus aureus* virulence factors in evasion from innate immune defenses in human and animal diseases. Immunol Lett. 2013;150:12–22. https://doi.org/10.1016/j.imlet.2013.01.004.
40. Liang X, et al. Inactivation of a two-component signal transduction system, SaeRS, eliminates adherence and attenuates virulence of *Staphylococcus aureus*. Infect Immun. 2006;74:4655–65. https://doi.org/10.1128/IAI.00322-06.
41. Javed MR, et al. CRISPR-Cas system: history and prospects as a genome editing tool in microorganisms. Curr Microbiol. 2018;75:1675–83. https://doi.org/10.1007/s00284-018-1547-4.
42. Guan J, Wang W, Sun B. Chromosomal targeting by the type III-A CRISPR-Cas system can reshape genomes in *Staphylococcus aureus*. mSphere. 2017. https://doi.org/10.1128/mSphere.00403-17.
43. Cao L, et al. Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of *Staphylococcus aureus*. Int J Med Microbiol. 2016;306:686–96. https://doi.org/10.1016/j.ijmm.2016.08.005.
44. Choueiry F, Xu R, Zhu J. Adaptive metabolism of *Staphylococcus aureus* revealed by untargeted metabolomics. J Proteome Res. 2022;21:470–81. https://doi.org/10.1021/acs.jproteome.1c00797.
45. Thomer L, et al. N-acetylglucosaminylation of serine-aspartate repeat proteins promotes *Staphylococcus aureus* bloodstream infection. J Biol Chem. 2014;289:3478–86. https://doi.org/10.1074/jbc.M113.532655.

Publisher's Note