



Identification of Three Clf-Sdr Subfamily Proteins in *Staphylococcus warneri*, and Comparative Genomics Analysis of a Locus Encoding CWA Proteins in *Staphylococcus* Species

OPEN ACCESS

Edited by:

Daniel Yero,
Autonomous University of Barcelona,
Spain

Reviewed by:

Akhilesh Kumar Chaurasia,
Sungkyunkwan University,
South Korea
Keenan A. Lacey,
New York University, United States

*Correspondence:

Mei Zhu
zhumei_d@163.com
Qiyu Bao
baoqy@genomics.cn
Hailin Zhang
zhlwz97@hotmail.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 05 April 2021

Accepted: 28 June 2021

Published: 29 July 2021

Citation:

Sun Z, Zhang X, Zhou D, Zhou K,
Li Q, Lin H, Lu W, Liu H, Lu J, Lin X,
Li K, Xu T, Zhu M, Bao Q and Zhang H
(2021) Identification of Three Clf-Sdr
Subfamily Proteins in *Staphylococcus
warneri*, and Comparative Genomics
Analysis of a Locus Encoding CWA
Proteins in *Staphylococcus* Species.
Front. Microbiol. 12:691087.
doi: 10.3389/fmicb.2021.691087

Zhewei Sun^{1,2,3†}, Xueya Zhang^{1,2,3†}, Danying Zhou^{1,3†}, Kexin Zhou^{1,3}, Qiaoling Li^{1,2,3},
Hailong Lin^{1,2,3}, Wei Lu^{1,3}, Hongmao Liu^{1,2,3}, Junwan Lu^{1,3}, Xi Lin^{1,3}, Kewei Li^{1,3}, Teng Xu⁴,
Mei Zhu^{5*}, Qiyu Bao^{1,3*} and Hailin Zhang^{2,3*}

¹ Key Laboratory of Medical Genetics of Zhejiang Province, Key Laboratory of Laboratory Medicine, Ministry of Education, School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou, China, ² The Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou, China, ³ School of Laboratory Medicine and Life Sciences, Institute of Biomedical Informatics, Wenzhou Medical University, Wenzhou, China, ⁴ Institute of Translational Medicine, Baotou Central Hospital, Baotou, China, ⁵ Department of Clinical Laboratory, Zhejiang Hospital, Hangzhou, China

Coagulase-negative *Staphylococcus warneri* is an opportunistic pathogen that is capable of causing several infections, especially in patients with indwelling medical devices. Here, we determined the complete genome sequence of a clinical *S. warneri* strain isolated from the blood culture of a 1-year-old nursing patient with acute upper respiratory infection. Genome-wide phylogenetic analysis confirmed the phylogenetic relationships between *S. warneri* and other *Staphylococcus* species. Using comparative genomics, we identified three cell wall-anchored (CWA) proteins at the same locus (*sdr*), named SdrJ, SdrK, and SdrL, on the chromosome sequences of different *S. warneri* strains. Structural predictions showed that SdrJ/K/L have structural features characteristic of Sdr proteins but exceptionally contained an unusual N-terminal repeat region. However, the C-terminal repetitive (R) region of SdrJ contains a significantly larger proportion of alanine (142/338, 42.01%) than the previously reported SdrI (37.00%). Investigation of the genetic organization revealed that the *sdrJ/K/L* genes were always followed by one or two glycosyltransferase genes, *gtfA* and *gtfB* and were present in an ~56 kb region bordered by a pair of 8 bp identical direct repeats, named Sw-Sdr. This region was further found to be located on a 160-kb region subtended by a pair of 160-bp direct repeats along with other virulence genes and resistance genes. Sw-Sdr contained a putative integrase that was probably a remnant of a functional integrase.

Evidence suggests that Sw-Sdr is improbably an efficient pathogenicity island. A large-scale investigation of *Staphylococcus* genomes showed that *sdr* loci were a potential hotspot of insertion sequences (ISs), which could lead to intraspecific diversity at these loci. Our work expanded the repository of *Staphylococcus* Sdr proteins, and for the first time, we established the connection between *sdr* loci and phylogenetic relationships and compared the *sdr* loci in different *Staphylococcus* species, which provided large insights into the genetic environment of CWA genes in *Staphylococcus*.

Keywords: cell wall anchored proteins, MSCRAMM, SdrJ/K/L, *Staphylococcus warneri*, comparative genomics, *sdr* loci

IMPORTANCE

Staphylococcus warneri can occasionally cause infection in patients with a compromised immune system. The emerging of multidrug-resistant (MDR) strains among *Staphylococcus*, *S. warneri* and other coagulase-negative *Staphylococcus* species has attracted the attention of scientists. *Staphylococcus* as gram-positive bacteria can express variety of cell wall-anchored (CWA) proteins that are considered as major virulence factors on their surface. Unfortunately, the genetic environment of these important virulence genes has not been systematically studied. In this study, we report three different *sdr* genes, encoding cell wall anchored proteins that belong to Clf-Sdr Subfamily, which locate at the same locus (*sdr*) of the chromosome sequences of different *S. warneri* strains. Comparison of the CWA genes at *sdr* loci in different *Staphylococcus* strains revealed both interspecific and intraspecific diversities at this locus. Importantly, we established the connection between the *sdr* locus and phylogenetic relationship and proposed that the *sdr* locus might be the potential hotspot of recombination, which would lead to intraspecific diversity at this locus. Our research provides insight into genetic environment of CWA genes in *Staphylococcus*, which fills in the gap of this field.

INTRODUCTION

Staphylococci continue to be the leading cause of many human infections. Along with *Staphylococcus aureus* and other coagulase-positive staphylococcal (CoPS) strains, whose pathogenicity and mechanisms of virulence are well elucidated, the coagulase negative species has recently garnered more attention. Albeit less virulent than coagulase-positive isolates (they express a restricted number of virulence factors), nowadays they are the most prevalent microorganisms within the genus, particularly in hospital settings, where iatrogenic procedures have improved patient care, but have also promoted body invasion by such pathogens.

Staphylococcus warneri, along with *Staphylococcus epidermidis*, is one of the predominant coagulase-negative staphylococcal (CoNS) bacteria inhabiting the skin in healthy humans (Nagase et al., 2002). In comparison with *S. aureus* and other CoPS strains that can cause a range of illnesses, from minor skin infections to life-threatening diseases, *S. warneri* prefers

to cause infections if the patients have underlying conditions such as indwelling foreign bodies and/or immunosuppression. As a common saprophyte of human epithelia, *S. warneri* is also frequently isolated from saliva, dental plaque, and nasal swabs, where it represents the third most prevalent CoNS species after *S. epidermidis* and *S. hominis* (Ohara-Nemoto et al., 2008). Over the past three decades, with the progressive refinement of identification techniques, *S. warneri* has been reported as a new pathogenic species frequently isolated from bacteremia, sepsis with multiple abscesses, orthopedic infections, vertebral osteomyelitis, and ventricular shunt infections of immunocompromised patients (Bryan et al., 1987; Kamath et al., 1992; Campoccia et al., 2010; Martínez-Lage et al., 2010). In addition, the overuse of antibiotics has led to a rapid increase in the occurrence of multiresistant strains of many bacteria, including *S. warneri*. Currently, the importance of *S. warneri* as a modern-day pathogen is growing, as it has established itself as a successful nosocomial pathogen.

Pathogenic gram-positive cocci express a plethora of virulence factors, including cell wall-anchored (CWA) proteins involved in microbial attachment to host tissues, which are considered the first critical step in the establishment of most bacterial infections (Foster and Höök, 1998). *S. aureus* can express up to 24 different CWA proteins, whereas CoNS species such as *S. epidermidis* and *S. lugdunensis* express a smaller number of CWA proteins (Bowden et al., 2005; Heilbronner et al., 2011). Timothy et al., proposed to classify *Staphylococcus* CWA proteins into four groups, namely, microbial surface component recognizing adhesive matrix molecule (MSCRAMM) family, near-iron transporter (NEAT) motif family, three-helical bundle family and G5-E repeat family, on the basis of the presence of motifs that have been defined by structure-function analysis (Foster et al., 2014). However, to the best of our knowledge, there are still numerous CWA proteins in *Staphylococcus* that cannot be assigned to any groups. The most prevalent group in *Staphylococcus* is the MSCRAMM family, which is defined by tandemly linked IgG-like folded domains (N2N3) that can engage in ligand binding by the dock, lock and latch (DLL) mechanism (Ponnuraj et al., 2003; Foster et al., 2014; Foster, 2019). The Clf-Sdr protein, with a repeat (R) region composed of serine-aspartate dipeptide repeats at the C-terminus, defines a subfamily of MSCRAMMs. The Clf-Sdr subfamily in *S. aureus* includes clumping factor A/B (ClfA/B) and SdrC/D/E (Foster et al., 2014), while in *S. epidermidis* it includes SdrG/F (Bowden et al., 2005). The N2N3 domains of these Sdr proteins share low sequence

identities, which allow them to have different functions and target different ligands or the same ligands at different sites (Foster and Höök, 1998; Wang et al., 2013). For example, ClfA and SdrG have been shown to only bind to fibrinogen molecules at different sites, while ClfB is reported to adhere to multiple peptides carrying the glycine-serine-rich (GSR) motif, such as fibrinogen α (Fg α), cytokeratin 10 (CK10) peptides and dermokines (Ganesh et al., 2011; Xiang et al., 2012). Unfortunately, the MSCRAMMs in CoNS, except for *S. epidermidis*, draw less attention to researchers; for example, MSCRAMMs from *S. warneri* have not been characterized yet.

Cell wall-anchored proteins can be encoded on mobile genetic elements (MGEs), such as biofilm-associated protein (Bap) (Ubeda et al., 2003), *S. epidermidis* surface protein J (SesJ) (Arora et al., 2020) and *S. saprophyticus* surface protein F (SssF) (King et al., 2012). The presence of virulence factors in MGEs can influence the virulence potential of a specific strain. Therefore, it is critical to study the genetic organization of virulence factors in emerging pathogenic bacteria to understand the mechanisms employed by these bacteria to cause infections. Sdr proteins, as an interesting subfamily of MSCRAMMs, have attracted much attention from researchers, but most studies have focused on their interaction with host proteins. Based on previous studies, we found that some of these Sdr proteins were obviously a part of accessory genes, such as SdrD/E from *S. aureus* (Liu et al., 2015). However, what leads to such a situation remains elusive.

Here, we preliminarily characterized three different Sdr proteins at the same locus in *S. warneri* strains and performed a comprehensive comparative genomics analysis of the locus in different *Staphylococcus* species.

MATERIALS AND METHODS

Bacterial Strains, Genome Sequencing, and Genome Assembly

Staphylococcus warneri WS479 was isolated from the blood culture of a 1-year-old nursing patient with acute upper respiratory infection in an affiliated hospital of Wenzhou Medical University, Zhejiang, China. The strain was identified by a Vitek-60 microorganism autoanalysis system (BioMerieux Corporate, Craponne, France).

The AxyPrep Bacterial Genomic DNA Miniprep Kit (Axygen Scientific, Union City, CA, United States) was used to extract the genomic DNA of *S. warneri* WS479. Library preparation, MinION sequencing and Illumina sequencing were performed at Shanghai Sunny Biotechnology Co., Ltd. The MinION long reads were initially assembled by Canu v1.8 (Koren et al., 2017), and then two FASTQ sequence files generated using the Illumina HiSeq 2500 platform were mapped onto the primary assembly to control assembly quality and to correct possible misidentified bases by using Bwa (Li and Durbin, 2009).

The genomes used to determine phylogenetic relationships and perform comparative genomics analysis were downloaded from the NCBI genome database. For those genomes used to investigate the *sdr* locus, the complete genome sequences were included preferentially, and the draft genome sequences

that incompletely harbored the *sdr* locus were excluded. All the genomes used in this study were confirmed by calculating average nucleotide identity (ANI) (Supplementary Figure 5) using FastANI (Jain et al., 2018).

Antimicrobial Susceptibility Testing and Cloning Experiments

The minimum inhibitory concentrations (MICs) were determined using the agar dilution method following the guidelines of the Clinical and Laboratory Standards Institute (CLSI, 2019). Susceptibility patterns were interpreted according to the CLSI breakpoint criteria for *Staphylococcus*. MIC was defined as the lowest concentration producing no visible bacterial growth. *Enterococcus faecalis* ATCC 29212 and *Escherichia coli* ATCC 25922 were used as reference strains for quality control. The resistance gene sequences (*aadD2* and *blaZ*) along with their promoter regions were PCR-amplified using the primers 5'-GCTCTAGAGCTTTCTATTATTGCAATGTGGAATTG-3' and 5'-CGGGATCCCGTCAAATGGTATGCGTTTGTACACA-3' for *aadD2* and 5'-CGGGATCCCGATTTAGCCATTTTGACACCTTCTTT-3' and 5'-CCAAGCTTGGTTAAAATTCCTTCAT TACTCTTTGGCG-3' for *blaZ*, with each having a pair of flanking restriction endonuclease adapters (*Xba*I and *Bam*HI for *aadD2* and *Bam*HI and *Hind*III for *blaZ*). The PCR products of *aadD2* and *blaZ* were then eluted from agarose gel, digested with the corresponding restriction endonucleases, and ligated into the pAM401 and pUCP24 vectors, respectively. The recombinant plasmid (pAM401-*aadD2*) was transformed into *E. faecalis* JH2-2 by the electroporation method, and the transformants were grown on brain heart infusion agar plates supplemented with chloramphenicol (16 μ g/mL). The recombinant plasmid (pUCP24-*blaZ*) was transformed into *E. coli* DH5 α cells via the calcium chloride method, and the transformants were grown on Luria-Bertani agar plates supplemented with gentamicin (40 μ g/mL). The recombinant plasmids were verified by both restriction endonuclease digestion and Sanger sequencing (Shanghai Sunny Biotechnology Co., Ltd., Shanghai, China).

Construction of a Phylogenetic Tree

To determine the phylogenetic relationship between *S. warneri* and other *Staphylococcus* species based on genomic data, we extracted and concatenated whole genome-wide single-copy orthologous protein sequences that were detected by OrthoFinder version 2.3.8 (Emms and Kelly, 2019) from 74 completely sequenced *Staphylococcus* genomes and a *Bacillus subtilis* genome (*Bacillus subtilis* 168, served as an outgroup). Multi-FASTA alignment was performed using MAFFT v7.407 (Katoh and Standley, 2013). Then, Gblocks (Castresana, 2000) was used to select conserved blocks of the resulting alignment. RAxML version 8.2.12 (Stamatakis, 2014) was used to infer the phylogeny by the maximum likelihood algorithm (ML) under the substitution matrix WAG which was chosen by ProtTest version 3.4 (Darriba et al., 2011).

Sequence Analysis and Molecular Modeling

Identification of domains and other sequence features that occurred within proteins was carried out using InterProScan (Jones et al., 2014), SignalP¹ and TMHMM². The repeat domains were identified using T-REKS³. Furthermore, several positively charged residues and LPXTG motifs at the C-terminus of CWA proteins were checked and confirmed manually. Comparison of the identities among proteins was performed using BLASTN and BLASTP (Altschul et al., 1990). Multiple sequence alignment was carried out using ClustalW⁴. The homology models of the N2N3 domains of SdrJ and SdrL were generated using the SWISS-MODEL repository using the crystal structure of the N2N3 domains of ClfB and SdrC (PDB IDs: 3AU0 and 6LXH) as the template, respectively. The homology model of the N2N3 domains of SdrK was ultimately established using MODELER (Webb and Sali, 2014) based on the crystal structures of SdrE and SdrG. Subsequently, the quality of the final models was assessed by a Ramachandran plot using the RAMPAGE server and quality model assessment with the protein structure analysis (ProSA) server (Wiederstein and Sippl, 2007; Kaufmann et al., 2011).

Protein Purification and Fluorescence Experiments

The DNA sequences of rSdrL178-748 (*S. warneri* WB224) and rSdrD54-680 (*S. aureus* TCH60) were synthesized with a pair of flanking restriction endonuclease adapters (*Nde*I for the forward primer and *Xho*I for the reverse primer) at Shanghai Tsingke Biotechnology Co., Ltd. Using pCold II as a cloning vector and *E. coli* BL21 as the recipient, transformants (pCold II-rSdr/BL21) were selected on LB agar plates containing 100 µg/mL ampicillin. The overnight culture of the recombinant strain (pCold II-rSdr/BL21) was diluted 100-fold in 100 ml of LB medium, incubated for 2–3 h at 37°C with orbital shaking at 250 rpm until the OD₆₀₀ reached 0.6, and incubated with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) (Sigma Chemicals Co., St. Louis, MO, United States) for 20 h at 16°C (Choi and Gelety, 2018). The recombinant protein was purified by affinity chromatography using BeyoGold His-tag Purification Resin (Beyotime, Shanghai, China) according to the manufacturer's instructions. The histidine tag was removed by Enterokinase (GenScript, Nanjing, China) for 16 h at 16°C.

Bis-ANS (4,4'-dianilino-1,1-binaphthyl-5,5'-disulfonic acid, dipotassium salt) emission spectra were measured using a fluorescence spectrophotometer (F-7000). rSdrL178-748/rSdrD54-680 (3 µM monomer) in 50 mM Tris-buffer and 10 mM NaCl (pH 7.5), was incubated for 1 h at 25°C in absence or presence of 1 mM bivalent CaCl₂ (Ca²⁺-bound) were saturated with 10 µM bis-ANS and their emission spectra were recorded from 430 to 600 nm using an excitation wavelength of 390 nm.

Gene Predictions, Genomic Islands Predictions, and Functional Annotations

Genes were predicted and preliminarily annotated by using Prokka v1.14.0 (Seemann, 2014). Prediction of genomic islands (GIs) was carried out by combining a host of tools,

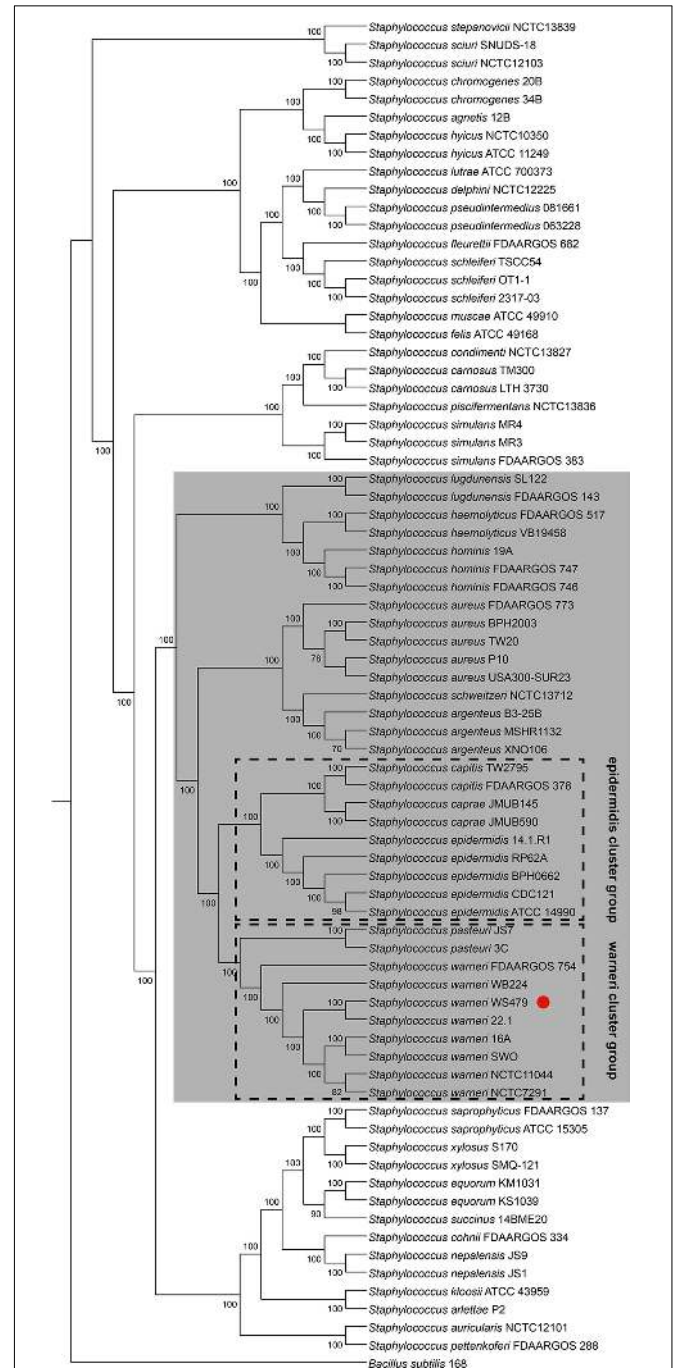


FIGURE 1 | Phylogenetic tree for 74 complete genome sequences composed of 38 *Staphylococcus* species. The interior values are the bootstrap probabilities after 1,000 replicates. The species belonging to group SD are shaded in gray. *S. warneri* WS479 is indicated with a red dot.

¹<http://www.cbs.dtu.dk/services/>

²www.cbs.dtu.dk/services/TMHMM-2.0/

³<https://bioinfo.crbm.cnrs.fr/index.php?route=tools&tool=3>

⁴<https://embnet.vital-it.ch/software/ClustalW.html>

such as Islandviewer 4 (Bertelli et al., 2017), PredictBias (Pundhir et al., 2008) and Alien Hunter (Vernikos and Parkhill, 2006). Direct repeats were identified using the inbuilt Find Repeat and Emboss tool fuzznuc in Geneious software. Easyfig (v2.2.3) software was used to generate the figures showing sequence comparisons and nucleotide identities of the *sdr* locus-containing region (Sullivan et al., 2011). Functional annotations were performed by querying different databases, such as the NCBI non-redundant (NCBI Resource Coordinators, 2018) and Swiss-Prot (UniProt Consortium, 2015) databases, using DIAMOND (Buchfink et al., 2015) with an evaluation threshold of $1e^{-5}$. Annotation of the resistance genes was performed using ResFinder (Zankari et al., 2012) and Resistance Gene Identifier (RGI) software version 4.0.3 of the Comprehensive Antibiotic Resistance Database (McArthur et al., 2013). Identification of virulence genes was carried out using BLASTP to search the protein sequences against the Virulence Factor Database (VFDB)⁵. Other bioinformatics tools were written using Python⁶.

⁵<http://www.mgc.ac.cn/VFs/>

⁶<https://www.python.org/>

Data Availability

The complete chromosome and two plasmid sequences (pWS-31 and pWS-25) of *S. warneri* WS479 have been submitted to DDBJ/EMBL/GenBank under accession numbers, CP061041.1, CP061042.1, and CP061043.1, respectively.

RESULTS AND DISCUSSION

General Features of the *S. warneri* WS479 Genome

The whole genome of *S. warneri* WS479 contained one single circular chromosome and two plasmids designated pWS-31 and pWS-25 (Supplementary Table 1). The chromosome is approximately 2.51 Mb in length (Supplementary Figure 1) and contains 2,491 ORFs, 5 rRNA-encoding gene (rDNA) clusters and 62 tRNA with an average GC content of 32.87%. The plasmids pWS-31 and pWS-25 have circularly closed DNA sequences of 31,000 bp and 25,160 bp in length with average GC contents of 29.86 and 30.46%, respectively.

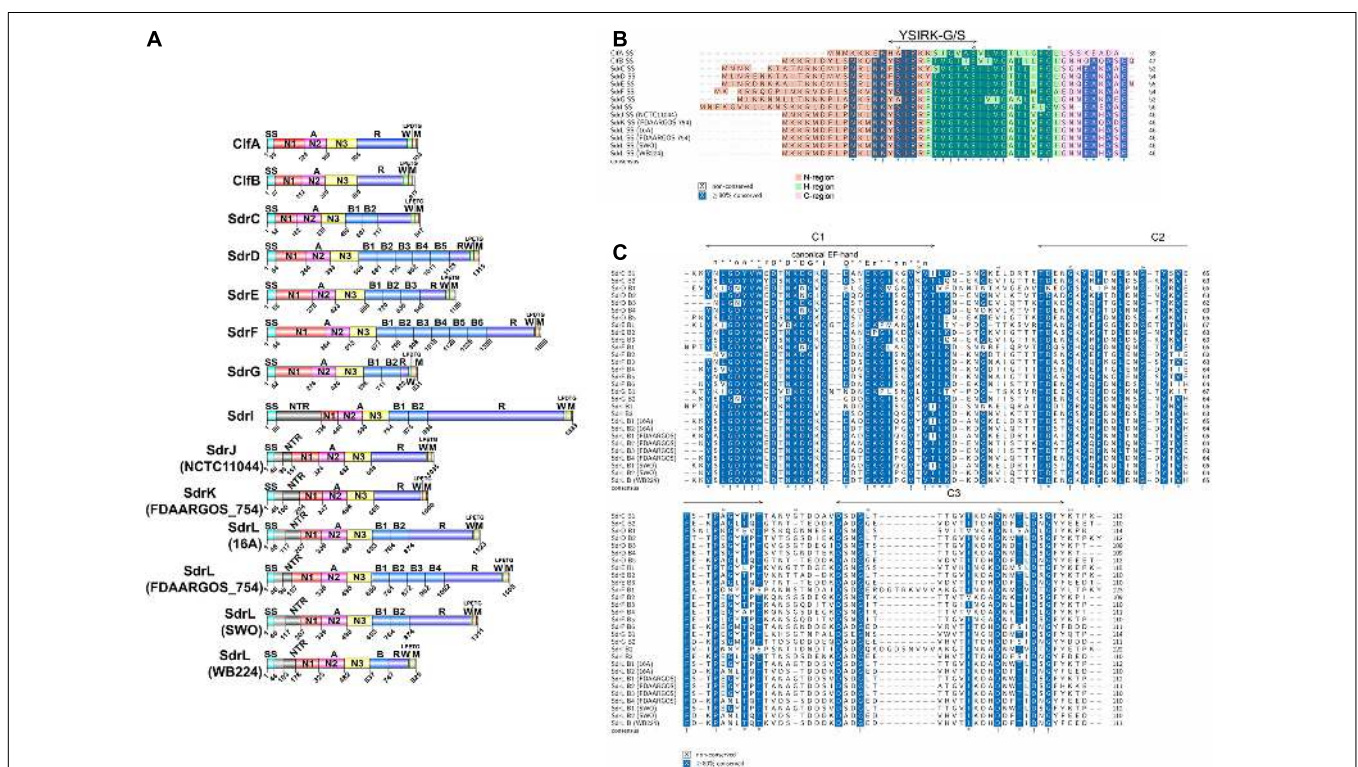
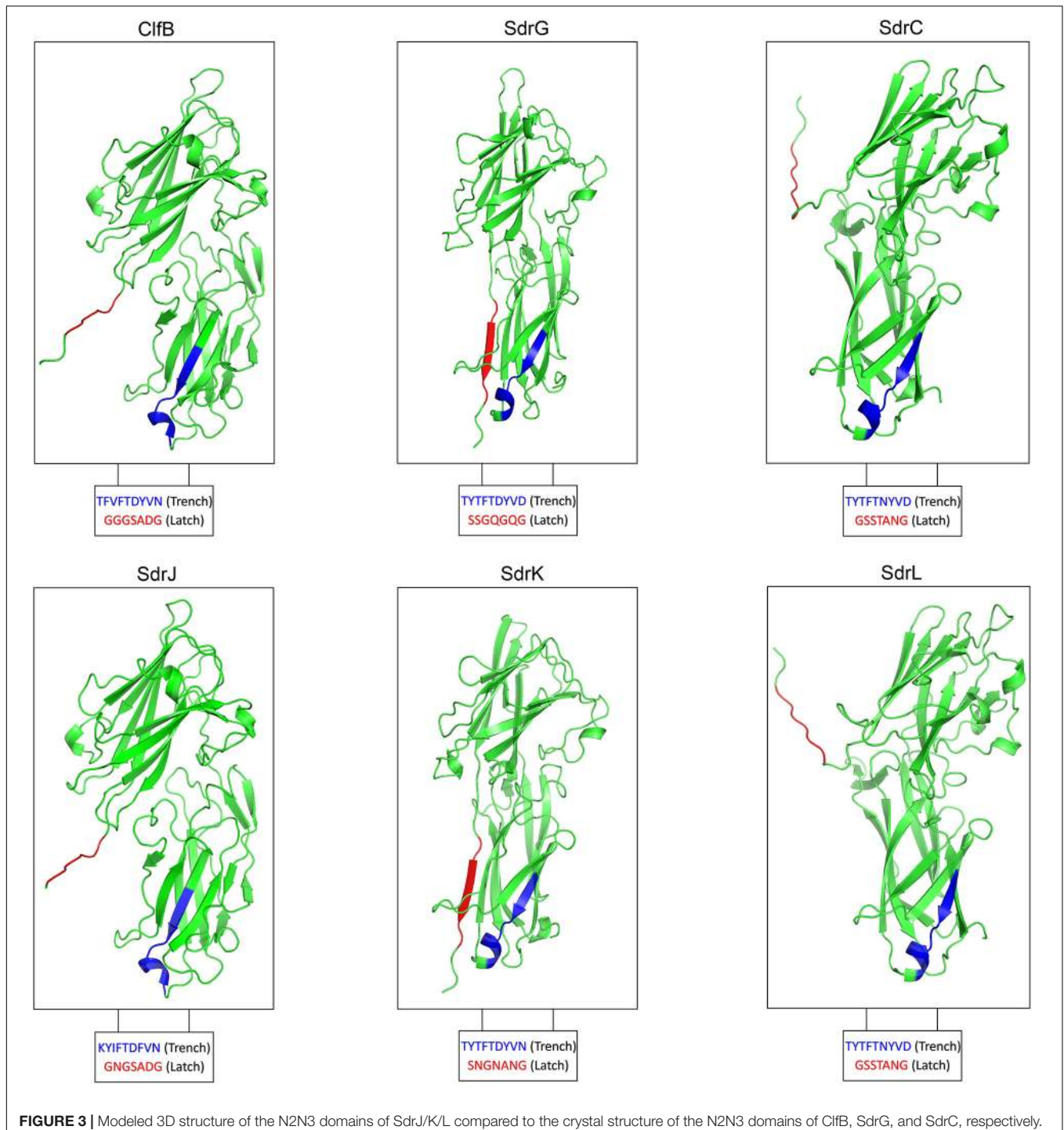


FIGURE 2 | Comparison of the SdrJ/K/L proteins (only those with complete genome sequences and complete *sdr* genes are included). **(A)** Complete schematic representation of the *S. warneri* SdrJ/K/L proteins and other known Sdr subfamily proteins. The signal sequence is shown in light blue, NTRs in dark gray, N1 domain in red, N2 domain in pink, N3 domain in yellow, B-repeats in light blue, R region in dark blue, and cell wall-spanning region (W) in green and membrane anchor (M) in orange. The LPXTG motif is indicated directly. Strain names are shown in brackets. The accession number of SdrJ from *S. warneri* NCTC11044 is VED76443.1; SdrL from *S. warneri* 16A is QDW97251.1; SdrL and SdrK from *S. warneri* FDAARGOS_754 are QKI08103.1 and QKI08104.1, respectively; SdrL from *S. warneri* SWO is WP_122135109.1; SdrL from *S. warneri* WB224 is QJX56829.1; the accession numbers of ClfA/B and SdrC/D/E/F/G/I are BAF67028.1, BAF68801.1, BAF66795.1, BAF66796.1, BAF66797.1, AIR82723.1, AAF72510.1, and AF402316.1, respectively. **(B)** Multiple sequence alignments of the signal peptides in Clf/Sdr proteins. N, H, C regions are shadowed in red, green and pink, respectively. **(C)** Multiple sequence alignments of the B-repeats in Sdr proteins. Each B-repeat contains three conserved stretches, namely, C1, C2, and C3, which are framed in green boxes. The consensus sequence for a canonical EF-hand is taken from that of Kretsinger (1987). "*" means $\geq 80\%$ conserved and "!" means 100% conserved.

The *in vitro* susceptibility testing of *S. warneri* WS479 exhibited resistance to a number of antibiotics (**Supplementary Table 2**), including erythromycin and clarithromycin (macrolides) and amikacin and azithromycin (aminoglycosides), according to CLSI breakpoint criteria for *Staphylococcus* (**Supplementary Table 2**). Moreover, the minimum inhibitory concentration (MIC) of roxithromycin against *S. warneri*

WS479 was 256 $\mu\text{g}/\text{mL}$, which was significantly higher than the resistance breakpoint for erythromycin ($>8 \mu\text{g}/\text{mL}$), though no interpretation criteria for the antimicrobial was available. Moreover, three antibiotic resistance genes (*blaZ*, *aadD2* and *msrA*) involved in drug resistance against three antibiotics (β -lactams, aminoglycosides and macrolides, respectively) were predicated on the plasmid pWS-31, and two of them (*blaZ*



and *aadD2*) were verified to be functional through cloning experiments (Supplementary Table 2).

Phylogenetic Relationship Between *S. warneri* and Other *Staphylococcus* Species

Lamers et al. (2012) established a new classification using 16S rRNA, *tuf*, *rpoB* and *dnaJ* genes to classify the *Staphylococcus* species into 15 cluster groups. *S. warneri* together with *S. pasteurii* were defined as the “warneri cluster group.” To confirm the phylogenetic relationship between *S. warneri* and other *Staphylococcus* species, we constructed a maximum likelihood phylogenetic tree based on concatenating deduced amino acid sequences of 622 single-copy ortholog sequences from 74 completely sequenced *Staphylococcus* genomes (including *S. warneri* WS479) and a *Bacillus subtilis* genome (*Bacillus subtilis* 168, served as an outgroup) available in NCBI. The results confirmed that *S. warneri* together with *S. pasteurii* fell into a single clade with perfect supporting bootstrap values (100%) (Figure 1). Furthermore, the “warneri cluster group” was clustered closest to the “epidermidis cluster group,” which was composed of *S. epidermidis*, *S. capitis*, and *S. caprae*, indicating similar biological traits among these species.

SdrJ/K/L Are Novel Clf-Sdr Subfamily Proteins

When searching for the putative CWA proteins in *S. warneri* genomes, we discovered three genes encoding proteins sharing similar structures with the Clf-Sdr family, which we have called SdrJ/K/L, according to the established nomenclature. The

deduced full-length SdrJ, SdrK and SdrL proteins in strains NCTC11044, FDAARGOS 754 and 16A are 1,035, 1,000, and 1,323 amino acids (aa) (Figure 2A) with predicted mature molecular masses of 100.56, 97.93, and 133.32 kDa, respectively. SdrJ/K/L are multidomain proteins that contain, starting from the N-terminus, a 46-amino acid long signal sequence, N-terminal repeats (NTRs), an A-region, 0–4 B-repeats, an SD (AD) repetitive (R) region and a typical cell wall anchoring sequence, such as an LPXTG motif, a hydrophobic membrane spanning region and a short cytoplasmic positively charged tail (Figure 2A). The A-region in these three proteins shares 35.90 to 45.36% sequence identity (Supplementary Table 3), suggesting that they have different functions. Surface proteins contain an N-terminal signal peptide that promotes their translocation across the bacterial membrane through the Sec pathway (Dramsai et al., 2008). The signal peptides of Sdr proteins are comprised a string of 17–18 hydrophobic aa (H-region), a variable-length N-region and an 8–10 aa C-region (Figure 2B). Multiple sequence alignments revealed that SdrJ/K/L like other reported Sdr proteins had a YSIRKXXXGXXS-like motif (Figure 2B). Deletion of this motif decreased the efficiency of signal peptidase, resulting in accumulation of precursors in the membrane and thus impacting on the secretion of CWA proteins (Dramsai et al., 2008).

The most prevalent CWA proteins are MSCRAMMs, which are defined by tandemly linked IgG-like folded domains that bind to their ligands through the “dock, lock, and latch” (DLL) mechanism. This binding mechanism involves characteristic structural features in the A-region, including two adjacent IgG-like folded domains with a conserved TYTFTDYVD-like motif presenting at the “back” of the latching trench in the first

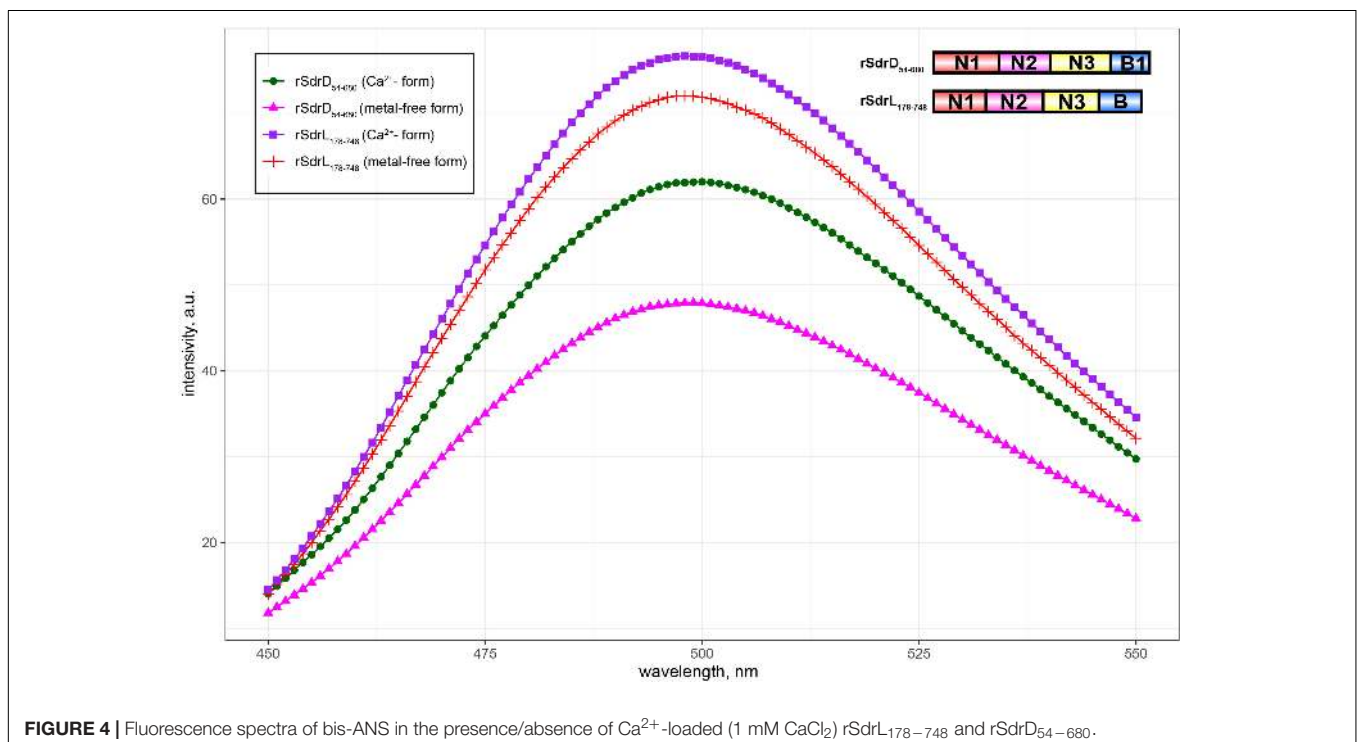


FIGURE 4 | Fluorescence spectra of bis-ANS in the presence/absence of Ca^{2+} -loaded (1 mM CaCl_2) rSdrL_{178–748} and rSdrD_{54–680}.

domain in the tandem and a latch sequence at the C-terminal extension of the second domain (Ponnuraj et al., 2003). A latch sequence is not a conserved sequence of amino acids but is formed by alternating small residues (Ponnuraj et al., 2003). Domain and tertiary structure prediction revealed that residues 324–649 of SdrJ, 347–665 of SdrK and 339–653 of SdrL adopt two IgG-like folds (N2N3) (Figures 2A, 3). They all furthermore contained a TYTFTDYVD-like motif at the expected positions for a latching trench in the first domain of the predicted IgG-like tandem and a putative latch sequence in the extension of the second IgG-like folded domain (Figure 3). The presence of a latch sequence and a conserved TYTFTDYVD-like motif in the N2N3 subdomain indicated that they could bind a ligand peptide by the DLL mechanism. Overall, the A-regions of these three novel Sdr proteins contain the characteristic IgG-like folded tandem of a MSCRAMM.

The A-region of SdrJ shared the highest sequence identity (48.04%) with that of ClfB. In addition, the B-repeat, a Ca²⁺-binding domain, was absent in both ClfB and SdrJ, which have very similar structures (Figure 2A). Significantly, however, the R region of SdrJ contained a larger proportion of alanine (42.01%), even than the previously reported SdrI (37.00%) (Sakinc et al., 2006). The R regions of Clf-Sdr proteins are hypothesized to extend the A-region away from the cell surface to prevent obstruction of ligand binding by the cell wall (Hartford et al., 1997; Lizcano et al., 2012); however, whether a rich of alanine acids in the R region of SdrJ and SdrI would influence ligand binding is still unclear.

The A-region of SdrK was similar to that of SdrG at the amino acid level (53.77%), suggesting that it may also have fibrinogen-binding activity (Davis et al., 2001; Sellman et al., 2008). B-repeats not only help project the A-region on the cell surface but also engage in ligand binding. For example, SdrF binds type I collagen via its B-repeats in a temperature-dependent manner (Di Poto et al., 2015). Thus, the absence of B-repeats in SdrJ and SdrK might weaken their ligand binding activity.

SdrL proteins are more prevalent than SdrJ and SdrK proteins in *S. warneri*. The A-region shared a significant sequence identity (73.50%) with that of SdrC, which was reported to involve β -neurexin binding as well as bacterial biofilm formation (Barbu et al., 2010, 2014). The B-repeats of *S. warneri* SdrL can vary from 1 to 4 copies. Sequence alignments from multiple B-repeats revealed that they also contained one conserved EF-hand motif (Figure 2C; Roman et al., 2016). To confirm the calcium binding ability of SdrL and investigate its change of hydrophobic surface in the presence of Ca²⁺, we purified rSdrL_{178–748} (A-region and one B-repeat) protein of *S. warneri* WB224 and rSdrD_{54–680} of *S. aureus* TCH60 (positive control). bis-ANS showed that compare with rSdrD_{54–680}, rSdrL_{178–748} had less decrease in fluorescence (14.23 a.u. versus 4.88 a.u.) when in the presence of Ca²⁺ at 500 nm (Figure 4), which suggested that the B repeat of SdrL was functional. In summary, SdrL may be related to biofilm formation, and its predominance indicates its importance to the pathogenicity of *S. warneri*.

Staphylococcal MSCRAMMs do not usually contain an NTR domain, which is found in these three proteins between the

signal sequence and the A-region (Arora et al., 2016). Unlike the NTR domains in SesJ and SdrI, the NTR domains found in the three *S. warneri* Sdr proteins were relatively short, and the tandem repeats were not identical (Table 1). As a result of the unusually short N1 region that connects the N2N3 tandem to the preceding NTR region, the A-region of these three proteins is smaller than that of their closest homologs. MSCRAMM family proteins bind to host proteins and can mediate bacterial adhesion, evade the immune response and participate in biofilm formation. Sequence analysis and modeling studies reveal that SdrJ/K/L harboring an unusual NTR domain are novel members of the *Staphylococcus* Clf-Sdr subfamily, which provides a point of penetration for further research on the pathogenicity of *S. warneri*. It is noteworthy that, *S. warneri* is the first CoNS possessing structural homologs (SdrJ/K) of Clf proteins that are free of B repeats. The interactions of SdrJ/K/L and their possible roles in *S. warneri* pathogenesis are currently being investigated.

Genetic Organization of the Region Encoding SdrJ/K/L in *S. warneri*

Comparative genomic analysis revealed that SdrJ/K/L were encoded on the same locus (hereafter named the *sdr* locus), which was located between *azo1* (encoding FMN-dependent NADPH-azoreductase) and *fole2* (encoding GTP cyclohydrolase) on the chromosome of *S. warneri*. Bioinformatic prediction

TABLE 1 | Summary of NTRs of the SdrJ/K/L proteins.

Protein (strain)	Sequence of NTR	Repeat number/Length of repeat sequence
SdrJ (NTCT11044)	EALTTAEEPKAEEKTK EAPATAEEPKAEEKTK EAPATAEEPKVDITTK EAPATAEKPADITTK	4/15
SdrL (16A)	TEGAPKADTTKEESA- TEEAPKADTTKEAPA- TEEAPKADTTKEAPA- TAEAPKADTTKEAPT- TEEAPKADTTKEAPVT TEEQTPKTTTDQAPE-	6/15
SdrL (FDAARGOS 754)	EASTTEKEPKADTTK EASTTEEAPKADTTK EAPTTAEELKVETTK EEPATAEKPADITTK	4/15
SdrL (SWO)	TEGAPKADTTKEESA- TEEAPKADTTKEAPA- TEEAPKADTTKEAPA- TAEAPKADTTKEAPT- TEEAPKADTTKEAPVT TEEQTPKTTTDQAPE-	6/15
SdrL (WB224)	-----EETPKADTTKEAPTA- -----EEPKADTTKEKATTK EAPTTAEEPADTTK----- EAPVTAEKPETDITTK----- EA-TTEEAPKVDITTK-----	5/15
SdrK (FDAARGOS 754)	-STTEEAPKADTTKEE -STTEEAPKADTTKEE PATTEEAPKADTTKE- ESVTEEAPKADATKE- ESATEEAPKADATKE- ESATEEAPKADTTKE- ESTTEEAPKADTTKE-	7/15

revealed that there were predicted genomic islands surrounding the locus encoding *SdrJ/K/L*. A search of direct repeat sequences resulted in finding an ~160 kb segment of the *S. warneri* chromosome, including the *sdrJ/K/L* genes flanked by 160 bp direct repeats (DRL1 and DRR1) (Figure 5 and Supplementary Figure 2). Approximately 40 kb upstream of the segment was a cluster of tRNA genes. Moreover, within the 160 kb segment, an ~56 kb subregion (hereafter named Sw-Sdr) was found to be bordered by 8 bp perfect direct repeats (Figure 5).

In total, we identified 9 or 10 putative virulence genes (depending on the strains) and two resistance genes in the 160 kb region (Figure 5). Notably, the *sdrJ/K/L* genes and their immediate surroundings within Sw-Sdr seemed to be poorly conserved. These *sdrJ/K/L* genes were always followed by one or two glycosyltransferases, *gtfA* and *gtfB* (Figure 5). Similar organization was also observed in other *Staphylococcus* species, such as downstream of the *sdrG* gene and *sdrC/D/E* genes in *S. epidermidis* and *S. aureus*, respectively (Hazebos et al., 2013). These *gtf* genes were proposed to glycosylate the R regions of Sdr proteins, which can protect Sdr proteins against host proteolytic activity and yet generate major epitopes for the human anti-staphylococcal antibody response (Hazebos et al., 2013). Thus, SdrJ/K/L might also subject to similar posttranslational modifications.

On the right side of Sw-Sdr was a putative tyrosine-integrase that was 185 aa in length. A domain search revealed that the integrase only possessed a catalytic domain and had several key residues substituted at the RHRV tetrad (Supplementary Figure 3; Kwon et al., 1997);

thus, it was an atypical integrase or might be a remnant of a functional integrase (Jo et al., 2016). Further BLAST analysis revealed that such atypical integrase seems to be frequently present in *Staphylococcus* genomes, but its detailed function remains unknown. Additionally, phylogenetic analysis revealed that these shortened integrases were phylogenetically distinct from known integrases of bacteriophages, integrons, integrative conjugative elements and other GIs (Supplementary Figure 4).

Overall, the genetic environment of *sdr* loci in *S. warneri* also harbored other virulence and resistance determinants; however, whether the region is a pathogenicity island remains unknown due to a lack of convincing evidence (this region has a similar GC content than the overall *S. warneri* genome (Supplementary Figure 1), is free of efficient mobility genes, and is free of genes from phages) (Hacker et al., 1997).

Not Only MSCRAMM Genes but Also Other CWA Genes Can Be Found at the *sdr* Locus

Different *S. warneri* strains possessed different *sdr* genes at the *sdr* locus (between *azo1* and *folE2*), suggesting that these *sdr* genes were a part of accessory genes or might be present in an MGE. After a large-scale investigation of *Staphylococcus* genomes, we found that species harbored CWA genes and (or) *gtf* genes at the *sdr* locus fell into a single clade (Figure 1), namely, group SD, while the remaining species contained transcriptional regulator or metabolism-related genes instead of CWA genes and (or) *gtf* genes (data not shown).

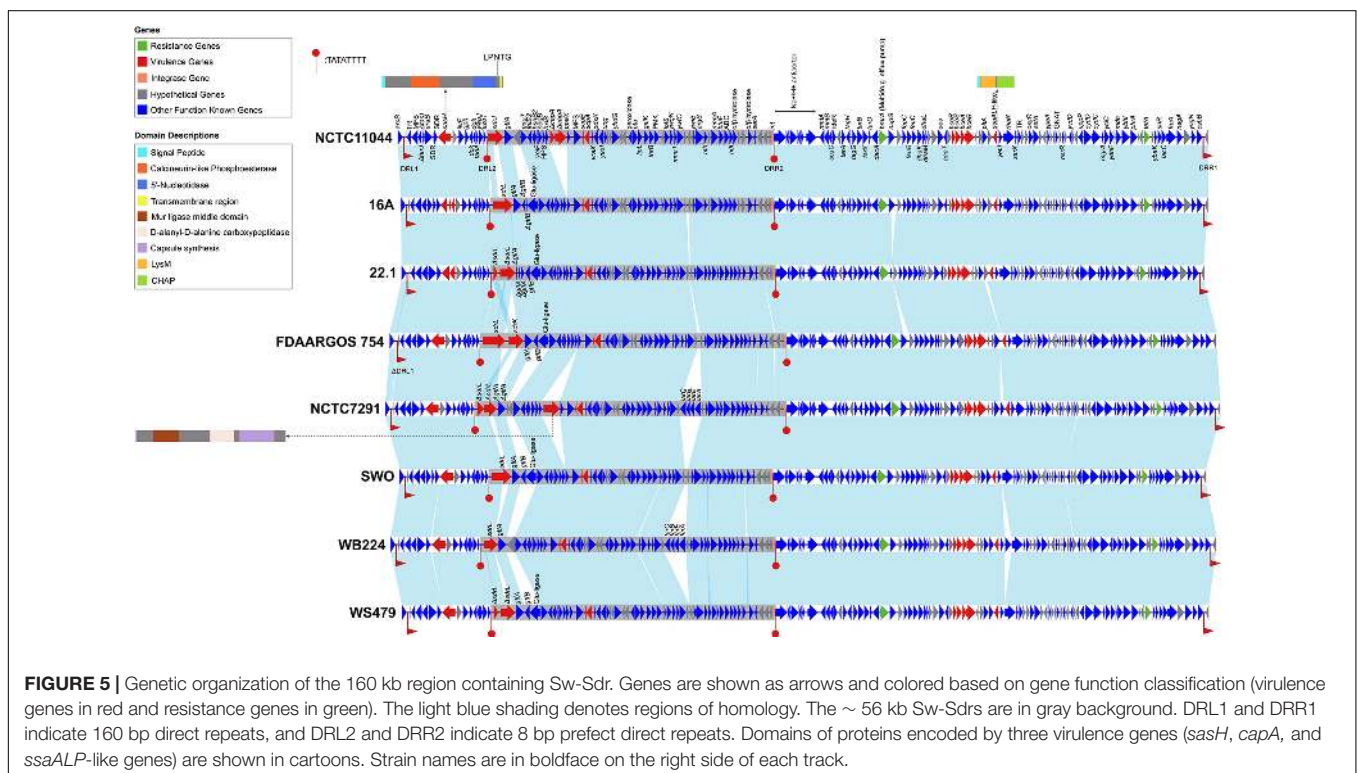


FIGURE 5 | Genetic organization of the 160 kb region containing Sw-Sdr. Genes are shown as arrows and colored based on gene function classification (virulence genes in red and resistance genes in green). The light blue shading denotes regions of homology. The ~56 kb Sw-Sdrs are in gray background. DRL1 and DRR1 indicate 160 bp direct repeats, and DRL2 and DRR2 indicate 8 bp perfect direct repeats. Domains of proteins encoded by three virulence genes (*sasH*, *capA*, and *ssaALP*-like genes) are shown in cartoons. Strain names are in boldface on the right side of each track.

To better characterize Sw-Sdr, we compared Sw-Sdr (*S. warneri* NCTC11044) with other relative regions from different *Staphylococcus* species. The results showed that except for the region containing CWA genes and their immediate surroundings, the other regions were relatively conserved among the investigated *Staphylococcus* species. In *S. aureus* and *S. epidermidis*, the *sdr* locus harbored the well-known Sdr family proteins SdrC/D/E and SdrG, respectively (Figure 6). In some *Staphylococcus* species, such as *S. capitis* and *S. caprae*, the *sdr* locus was free of CWA genes but still possessed the *gtf* gene. In *S. haemolyticus*, CWA genes at the *sdr* locus could encode not only MSCRAMM proteins but also G5-E repeat family proteins and Rib domain-containing proteins (Figure 6 and Table 2). Rib domain-containing proteins could also be found at the *sdr* loci in *S. hominis* and *S. lugdunensis*. Such proteins feature a C-terminal of tandem repeated Rib domains that are thought to define protective epitopes and may play a role in generating phenotypic and genotypic variation in group B *Streptococcus* (Michel et al., 1992; Wästfelt et al., 1996). These

facts suggest that the downstream integrase was independent of the diversity of CWA genes at the *sdr* locus because in *S. epidermidis*, the downstream integrase was absent, while in *S. lugdunensis*, the *sdr* locus and the integrase-containing region were not continuous. The integrase might be associated with the initial integration of this region in *Staphylococcus* but lost its function during evolution, thereby leading to the settling of this region.

Insertion Sequences at *sdr* Locus

We noticed that within a *Staphylococcus* species, although the CWA genes at this locus varied from strain to strain, the repository of these CWA genes was settled. To prove this, we investigated the genome sequences of the 11 *Staphylococcus* species belonging to group SD (Figure 1). All the investigated species showed a settled repository of CWA genes at this locus, such as *sdrC/D/E* in *S. aureus*, *sdrG* in *S. epidermidis*, and *sdrJ/K/L* in *S. warneri* and a gene encoding the Rib domain-containing protein in *S. lugdunensis* (Table 2). Moreover, species belonging

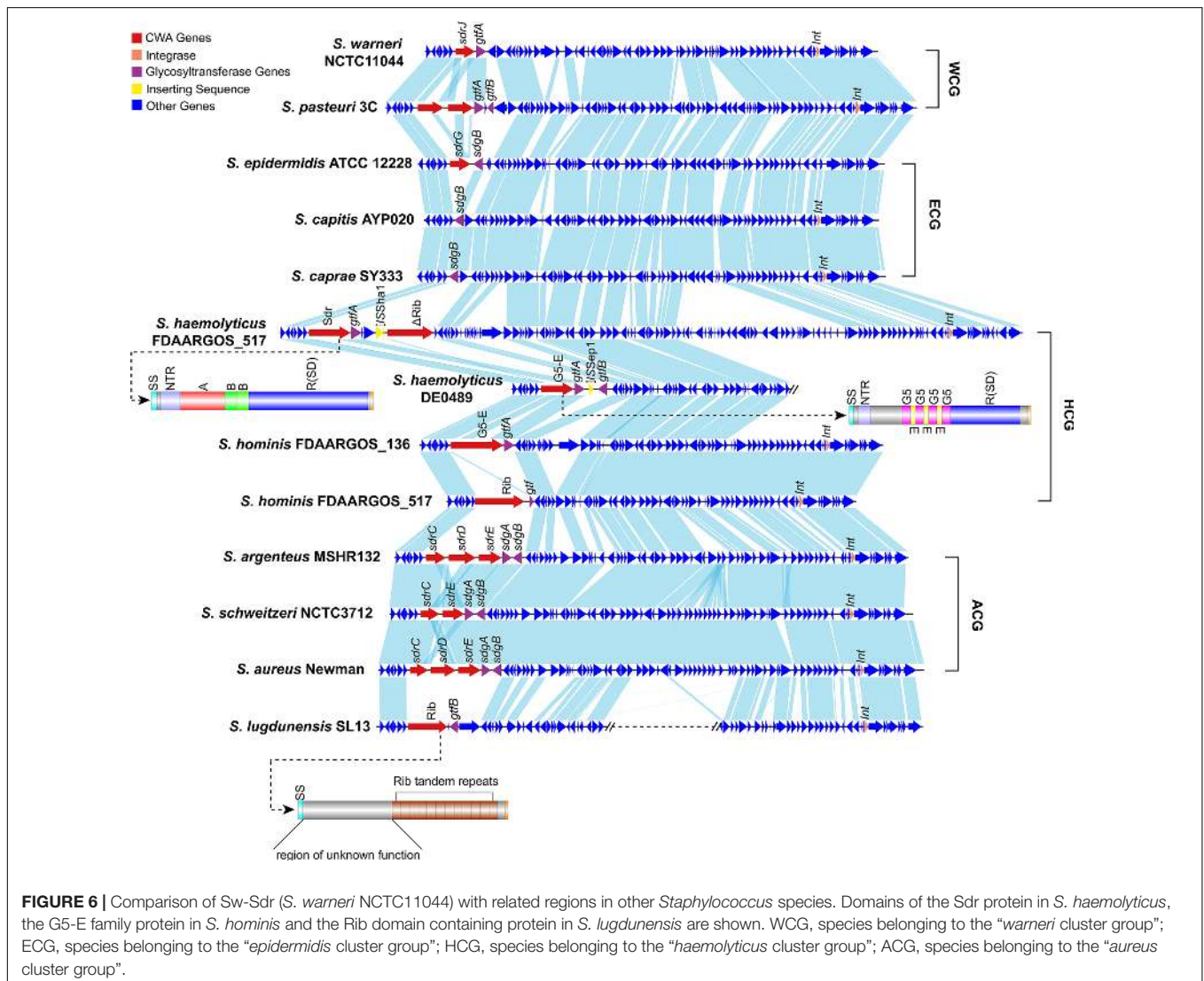


TABLE 2 | Statistics of CWA genes and ISs at *sdr* loci in 11 *Staphylococcus* species belong to group SD.

Species (Complete/Draft)	<i>S. aureus</i> (150/0)	<i>S. argenteus</i> (6/33)	<i>S. schweitzeri</i> (0/4)	<i>S. epidermidis</i> (45/39)	<i>S. caprae</i> (n = 5/0)/ <i>S. capitis</i> (4/22)	<i>S. haemolyticus</i> (n = 14/27)	<i>S. hominis</i> (12/11)	<i>S. lugdunensis</i> (25/5)	<i>S. warneri</i> (8/27)	<i>S. pasteurii</i> (2/5)
CWA proteins	SdrC (99.33%; n = 149) SdrD (92%; n = 138) SdrE (90%; n = 135)	SdrC-like (92.31%; n = 36) SdrD-like (79.49%; n = 31) SdrE-like (79.49%; n = 31)	SdrC-like (100.00%; n = 4) SdrD-like (25.00%; n = 1) SdrE-like (100.00%; n = 4)	SdrG (86.9%; n = 73)	NA	Rib domain containing (29.27%; n = 12) G5-E repeat family (31.71%; n = 13) Sdr family (56.10%; n = 23)	Rib domain containing (60.87%; n = 14) G5-E repeat family (39.13%; n = 9)	Rib domain containing (100.00%; n = 30)	SdrJ (8.57%; n = 3) SdrK (5.71%; n = 2) SdrL (88.57%; n = 31)	SdrL-like (85.71%; n = 6) SdrM (57.14%; n = 4)
CWA proteins positive rate	100.00%; n = 150	100.00%; n = 39	100%; n = 4	86.9%; n = 73	0.00%; n = 0	90.24%; n = 37	91.30%; n = 21	100.00%; n = 30	100%; n = 35	85.71%; n = 6
ISs	IS30 family (5.33%; n = 8) ISL3 family (1.33%; n = 2) IS256 family (2.00%; n = 3) IS1182 family (1.33%; n = 2)	NA	NA	ISSep3 (2.38%; n = 2)	NA	IS256 family (4.88%; n = 2) IS1182 family (39.02%; n = 16) ISL3 family (31.71%; n = 13)	IS1182 family (4.35%; n = 1)	NA	NA	NA
ISs positive rate	10.00%; n = 15	0.00%; n = 0	0.00%; n = 0	2.38%; n = 2	0.00%; n = 0	58.54%; n = 24	4.35%; n = 1	0.00%; n = 0	0.00%; n = 0	0.00%; n = 0

For CWA genes in *S. haemolyticus*, *S. hominis* and *S. lugdunensis*, CWA genes are grouped based on presence key domains (N2N3 domains, B-repeats, G5-E domain, and Rib domain), and for other species, CWA genes are grouped based on sequence identities (Supplementary Figure 6).

to the same cluster group tended to possess similar CWA genes at the *sdr* locus; for example, the *sdrC/D/E* derivatives in *S. argenteus* and *S. schweitzeri* share high sequence identities (>80%) with *sdrC/D/E* genes in *S. aureus*.

We serendipitously observed that some *Staphylococcus* genomes occasionally integrated an insertion sequence (IS) at the *sdr* locus. More specifically, we noticed that 10.00% (15/150) of *S. aureus* strains possessed IS at this locus, probably due to a lack of complete genome sequences; only 2.38% (2/84) of *S. epidermidis* strains and 4.35% (1/23) of *S. hominis* strains were found to carry IS (Table 2). Complete genome sequence data are essential for the analysis of genes containing repeat regions, such as the *sdr* locus. Unfortunately, due to lack of genome resources, the IS element was not observed at the *sdr* locus in some *Staphylococcus* species. *S. haemolyticus* was a special case with more than half (58.54%; 24/41) of the investigated genomes possessing the IS element at the *sdr* locus; of note, 78.57% (11/14) of the completely sequenced strains possessed the IS element at the *sdr* locus, implying that this number would be higher if more complete genome sequences were included. IS element was reported to involve in the modulation of expression of virulence factors in many bacterial species including *Staphylococcus* (Ziebuhr et al., 1999; Rajeshwari and Sonti, 2000; Conlon et al., 2004; Simser et al., 2005; Perez et al., 2015). In our dataset, *S. lugdunensis* was the only species with a relatively large number of complete genomes that no IS element was identified at *sdr* loci. Further investigation revealed that the Rib domain-containing protein in *S. lugdunensis* was not similar with other Sdr or G5-E proteins that possessed R region. Thus, the repeat region of Sdr or G5-E proteins at *sdr* loci could act as a recombination hotspot for the integration of such mobile genetic elements. However, further experiments will be needed to confirm this hypothesis. Therefore, we thought that the intraspecific differences in CWA genes or truncation of CWA genes (such as *sdrL* in *S. warneri* NCTC7291 and *S. warneri* WS479) at this locus were caused by these mobile genetic elements.

However, it is still unclear whether the putative site-specific integrase downstream of the *sdr* locus is involved in the formation of different categories of CWA proteins in various *Staphylococcus* species. Since *sdr* loci variably contained CWA genes within a *Staphylococcus* species, it can be a potential locus used for typing strains with different pathogenicity.

CONCLUSION

In this study, we determined the complete whole-genome sequence of *S. warneri* WS479, and confirmed that *blaZ* and *aadD2* on pWS-31 were functional. The phylogenetic relationship between *S. warneri* and other *Staphylococcus* species was confirmed by a genome-wide phylogenetic tree, and *S. warneri* together with *S. pasteurii* were in a single clade. Combining sequence analysis and modeling studies, we identified three CWA proteins, namely, SdrJ/K/L, on the chromosomes of *S. warneri*. These proteins have structural features characteristic of Sdr proteins but also contain an unusual N-terminal repeat

region. Further investigation of the genetic organization of *sdrJ/K/L* revealed that they were actually located at the same locus (*sdr* locus) on the chromosome sequence. Although a putative integrase and direct repeats were observed, our analysis still suggested that the region containing Sdr proteins (Sw-Sdr) was unlikely to be an efficient pathogenicity island. Comparative genomic analysis showed that Sw-Sdr was a conserved region found in different *Staphylococcus* species, and the MSCRAMM gene was not the only CWA gene at *sdr* loci. A large-scale investigation of *Staphylococcus* genomes showed that *sdr* loci were a potential hotspot of insertion sequences (ISs), leading to intraspecies diversity at these loci.

DATA AVAILABILITY STATEMENT

The complete chromosome and two plasmid sequences (pWS-31 and pWS-25) of *S. warneri* WS479 have been submitted to DDBJ/EMBL/GenBank under accession numbers, CP061041.1, CP061042.1, and CP061043.1, respectively.

ETHICS STATEMENT

This study was approved by The Ethics Committee of the Affiliated Hospital of Wenzhou Medical University (China) and informed consent was obtained from the patient.

AUTHOR CONTRIBUTIONS

ZS conceived the project and wrote the manuscript. DZ, XZ, and KZ performed the experiments. TX, QB, HZ, and MZ analyzed the results. All authors were involved in the design of the experiments and reviewed the manuscript.

FUNDING

This study was supported by the Natural Science Foundation of Zhejiang Province, China (LY19C060002 and LQ17H190001), Science and Technology Project of Wenzhou City, China (Y20170205 and 2019Y0358), National Natural Science Foundation of China (81973382 and 81960381), and Science

and Technology Project of Inner Mongolia Autonomous Region, China (201802125).

ACKNOWLEDGMENTS

The authors thank all the colleagues and reviewers who helped this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.691087/full#supplementary-material>

Supplementary Figure 1 | Chromosomal comparison of *S. warneri* WS479 with 7 other *S. warneri* chromosome sequences. Strain names are shown in the legend. The innermost circle shows the scale in Mb.

Supplementary Figure 2 | Multiple sequence alignments of the 160 bp DRL1s/DRR1s.

Supplementary Figure 3 | Multiple sequence alignments of the putative integrases on Sw-Sdr in *S. warneri* or Sw-Sdr-related regions in other *Staphylococcus* species with Cre, XerD, and phage integrase. The RHRV tetrad is shaded in gray.

Supplementary Figure 4 | The inferred phylogenetic relationship of the putative integrase on Sw-Sdr in *S. warneri* or Sw-Sdr related regions in other *Staphylococcus* species belong to group SD in relation to representative tyrosine recombinases of the shufflon-specific DNA recombinase (Rci), *dusA*-associated integrases (DAI), phage integrase (Phage), integrative and conjugative element (IntC), *Staphylococcus* pathogenicity island (SPI), *Staphylococcus* resistance island (SRI), integron integrase (IntI), site-specific recombinase (XerDC), phage P1-like recombinases (Cre) and genomic island integrase (IntG) families. Protein accession numbers of sequences used to generate this phylogenetic tree are in brackets. The interior values are the bootstrap probabilities after 1,000 replicates.

Supplementary Figure 5 | Average nucleotide identity between each genome used to investigate the *sdr* locus.

Supplementary Figure 6 | Sequence identity comparison of CWA proteins (identified based on the presence of key domains) encoded at *sdr* loci in *Staphylococcus* species belonging to group SD.

Supplementary Table 1 | Genome features of *S. warneri* WS479.

Supplementary Table 2 | MIC values of antibacterial drugs for all strains ($\mu\text{g/mL}$).

Supplementary Table 3 | Sequence identity matrix table containing all the SDR subfamily proteins compared in terms of SS, A, B, R, W, and M regions.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Arora, S., Li, X., Hillhouse, A., Konganti, K., Little, S. V., Lawhon, S. D., et al. (2020). *Staphylococcus epidermidis* MSCRAMM SesJ is encoded in composite islands. *mBio* 11:e2911–19. doi: 10.1128/mBio.02911-19
- Arora, S., Uhlemann, A.-C., Lowy, F. D., and Hook, M. (2016). A novel MSCRAMM subfamily in coagulase negative *Staphylococcal* species. *Front. Microbiol.* 7:540. doi: 10.3389/fmicb.2016.00540
- Barbu, E. M., Ganesh, V. K., Gurusiddappa, S., Mackenzie, R. C., Foster, T. J., Sudhof, T. C., et al. (2010). beta-Neurexin is a ligand for the *Staphylococcus aureus* MSCRAMM SdrC. *PLoS Pathog.* 6:e1000726. doi: 10.1371/journal.ppat.1000726
- Barbu, E. M., Mackenzie, C., Foster, T. J., and Höök, M. (2014). SdrC induces *Staphylococcal* biofilm formation through a homophilic interaction. *Mol. Microbiol.* 94, 172–185. doi: 10.1111/mmi.12750
- Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoard, G., Winsor, G. L., et al. (2017). Island viewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343
- Bowden, M. G., Chen, W., Singvall, J., Xu, Y., Peacock, S. J., Valtulina, V., et al. (2005). Identification and preliminary characterization of cell-wall-anchored proteins of *Staphylococcus epidermidis*. *Microbiology* 151, 1453–1464. doi: 10.1099/mic.0.27534-0

- Bryan, C. S., Parisi, J. T., and Strike, D. G. (1987). Vertebral osteomyelitis due to *Staphylococcus warneri* attributed to a Hickman catheter. *Diagn. Microbiol. Infect. Dis.* 8, 57–59. doi: 10.1016/0732-8893(87)90048-4
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Campoccia, D., Montanaro, L., Visai, L., Corazzari, T., Poggio, C., Pegreff, F., et al. (2010). Characterization of 26 *Staphylococcus warneri* isolates from orthopedic infections. *Int. J. Artif. Organs* 33, 575–581. doi: 10.1177/039139881003300903
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Choi, T.-J., and Geletu, T. T. (2018). High level expression and purification of recombinant flounder growth hormone in *E. coli*. *J. Genet. Engin. Biotechnol.* 16, 347–355. doi: 10.1016/j.jgeb.2018.03.006
- CLSI. (2019). *Performance Standards for Antimicrobial Susceptibility Testing. CLSI supplement M100*. 29th Edn. Wayne, PA: Clinical and Laboratory Standards Institute.
- Conlon, K. M., Humphreys, H., and O’Gara, J. P. (2004). Inactivations of rsbU and sarA by IS256 represent novel mechanisms of biofilm phenotypic variation in *Staphylococcus epidermidis*. *J. Bacteriol.* 186, 6208–6219. doi: 10.1128/JB.186.18.6208-6219.2004
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). Prot test 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
- Davis, S. L., Gurusiddappa, S., McCrea, K. W., Perkins, S., and Höök, M. (2001). SdrG, a fibrinogen-binding bacterial adhesin of the microbial surface components recognizing adhesive matrix molecules subfamily from *Staphylococcus epidermidis*, targets the thrombin cleavage site in the Bbeta chain. *J. Biol. Chem.* 276, 27799–27805. doi: 10.1074/jbc.M103873200
- Di Poto, A., Papi, M., Trivedi, S., Maiorana, A., Gavazzo, P., Vassalli, M., et al. (2015). In vitro effect of temperature on the conformational structure and collagen binding of SdrF, a *Staphylococcus epidermidis* adhesin. *Appl. Microbiol. Biotechnol.* 99, 5593–5603. doi: 10.1007/s00253-015-6456-x
- Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008). Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol. Rev.* 32, 307–320. doi: 10.1111/j.1574-6976.2008.00102.x
- Emms, D. M., and Kelly, S. (2019). Ortho finder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Foster, T. J. (2019). The MSCRAMM family of cell-wall-anchored surface proteins of gram-positive cocci. *Trends Microbiol.* 27, 927–941. doi: 10.1016/j.tim.2019.06.007
- Foster, T. J., Geoghegan, J. A., Ganesh, V. K., and Höök, M. (2014). Adhesion, invasion and evasion: The many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* 12, 49–62. doi: 10.1038/nrmicro3161
- Foster, T. J., and Höök, M. (1998). Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol.* 6, 484–488. doi: 10.1016/s0966-842x(98)01400-0
- Ganesh, V. K., Barbu, E. M., Deivanayagam, C. C. S., Le, B., Anderson, A. S., Matsuka, Y. V., et al. (2011). Structural and biochemical characterization of *Staphylococcus aureus* clumping factor B/ligand interactions. *J. Biol. Chem.* 286, 25963–25972. doi: 10.1074/jbc.M110.217414
- Hacker, J., Blum-Oehler, G., Mühldorfer, I., and Tschäpe, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23, 1089–1097. doi: 10.1046/j.1365-2958.1997.3101672.x
- Hartford, O., Francois, P., Vaudaux, P., and Foster, T. J. (1997). The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. *Mol. Microbiol.* 25, 1065–1076. doi: 10.1046/j.1365-2958.1997.5291896.x
- Hazenbos, W. L. W., Kajihara, K. K., Vandlen, R., Morisaki, J. H., Lehar, S. M., Kwakkenbos, M. J., et al. (2013). Novel *Staphylococcal* glycosyltransferases SdgA and SdgB mediate immunogenicity and protection of virulence-associated cell wall proteins. *PLoS Pathog.* 9:e1003653. doi: 10.1371/journal.ppat.1003653
- Heilbronner, S., Holden, M. T. G., van Tonder, A., Geoghegan, J. A., Foster, T. J., Parkhill, J., et al. (2011). Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors. *FEMS Microbiol. Lett.* 322, 60–67. doi: 10.1111/j.1574-6968.2011.02339.x
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- Jo, C. H., Kim, J., Han, A.-R., Park, S. Y., Hwang, K. Y., and Nam, K. H. (2016). Crystal structure of Thermoplasma acidophilum XerA recombinase shows large C-shape clamp conformation and cis-cleavage mode for nucleophilic tyrosine. *FEBS Lett.* 590, 848–856. doi: 10.1002/1873-3468.12109
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kamath, U., Singer, C., and Isenberg, H. D. (1992). Clinical significance of *Staphylococcus warneri* bacteremia. *J. Clin. Microbiol.* 30, 261–264. doi: 10.1128/JCM.30.2.261-264.1992
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kaufmann, K., Shen, N., Mizoue, L., and Meiler, J. (2011). A physical model for PDZ-domain/peptide interactions. *J. Mol. Model.* 17, 315–324. doi: 10.1007/s00894-010-0725-5
- King, N. P., Sakiñç, T., Ben Zakour, N. L., Totsika, M., Heras, B., Simerska, P., et al. (2012). Characterisation of a cell wall-anchored protein of *Staphylococcus saprophyticus* associated with linoleic acid resistance. *BMC Microbiol.* 12:8. doi: 10.1186/1471-2180-12-8
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Kretsinger, R. H. (1987). Calcium coordination and the calmodulin fold: divergent versus convergent evolution. *Cold Spring Harb. Symp. Quant. Biol.* 52, 499–510. doi: 10.1101/sqb.1987.052.01.057
- Kwon, H. J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276, 126–131. doi: 10.1126/science.276.5309.126
- Lamers, R. P., Muthukrishnan, G., Castoe, T. A., Tafur, S., Cole, A. M., and Parkinson, C. L. (2012). Phylogenetic relationships among *Staphylococcus* species and refinement of cluster groups based on multilocus data. *BMC Evol. Biol.* 12:171. doi: 10.1186/1471-2148-12-171
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Liu, H., Lv, J., Qi, X., Ding, Y., Li, D., Hu, L., et al. (2015). The carriage of the serine-aspartate repeat protein-encoding sdr genes among *Staphylococcus aureus* lineages. *Brazil. J. Infect. Dis.* 19, 498–502. doi: 10.1016/j.bjid.2015.07.003
- Lizcano, A., Sanchez, C. J., and Orihuela, C. J. (2012). A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. *Mol. Oral Microbiol.* 27, 257–269. doi: 10.1111/j.2041-1014.2012.00653.x
- Martínez-Lage, J. F., Martínez-Lage Azorín, L., and Almagro, M.-J. (2010). *Staphylococcus warneri* ventriculoperitoneal shunt infection: failure of diagnosis by ventricular CSF sampling. *Childs Nerv. Syst.* 26, 1795–1798. doi: 10.1007/s00381-009-1050-0
- McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13
- Michel, J. L., Madoff, L. C., Olson, K., Kling, D. E., Kasper, D. L., and Ausubel, F. M. (1992). Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci. *Proc. Natl. Acad. Sci. U S A.* 89, 10060–10064. doi: 10.1073/pnas.89.21.10060
- Nagase, N., Sasaki, A., Yamashita, K., Shimizu, A., Wakita, Y., Kitai, S., et al. (2002). Isolation and species distribution of *Staphylococci* from animal and human skin. *J. Vet. Res. Sci.* 64, 245–250. doi: 10.1292/jvms.64.245
- NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 46, D8–D13. doi: 10.1093/nar/gkx1095

- Ohara-Nemoto, Y., Haraga, H., Kimura, S., and Nemoto, T. K. (2008). Occurrence of *Staphylococci* in the oral cavities of healthy adults and nasal oral trafficking of the bacteria. *J. Med. Microbiol.* 57, 95–99. doi: 10.1099/jmm.0.47561-0
- Perez, M., Calles-Enriquez, M., del Rio, B., Ladero, V., Martín, M. C., Fernández, M., et al. (2015). IS256 abolishes gelatinase activity and biofilm formation in a mutant of the nosocomial pathogen *Enterococcus faecalis* V583. *Can. J. Microbiol.* 61, 517–519. doi: 10.1139/cjm-2015-0090
- Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., et al. (2003). A “dock, lock, and latch” structural model for a *Staphylococcal* adhesin binding to fibrinogen. *Cell* 115, 217–228. doi: 10.1016/s0092-8674(03)00809-2
- Pundhir, S., Vijayvargiya, H., and Kumar, A. (2008). PredictBias: A server for the identification of genomic and pathogenicity islands in prokaryotes. *Silico Biol.* 8, 223–234.
- Rajeshwari, R., and Sonti, R. V. (2000). Stationary-phase variation due to transposition of novel insertion elements in *Xanthomonas oryzae* pv. *oryzae*. *J. Bacteriol.* 182, 4797–4802. doi: 10.1128/JB.182.17.4797-4802.2000
- Roman, A. Y., Devred, F., Lobatchov, V. M., Makarov, A. A., Peyrot, V., Kubatiev, A. A., et al. (2016). Sequential binding of calcium ions to the B-repeat domain of SdrD from *Staphylococcus aureus*. *Can. J. Microbiol.* 62, 123–129. doi: 10.1139/cjm-2015-0580
- Sakinc, T., Kleine, B., and Gatermann, S. G. (2006). SdrI, a serine-aspartate repeat protein identified in *Staphylococcus saprophyticus* strain 7108, is a collagen-binding protein. *Infect. Immun.* 74, 4615–4623. doi: 10.1128/IAI.01885-05
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sellman, B. R., Timofeyeva, Y., Nanra, J., Scott, A., Fulginiti, J. P., Matsuka, Y. V., et al. (2008). Expression of *Staphylococcus epidermidis* SdrG increases following exposure to an in vivo environment. *Infect. Immun.* 76, 2950–2957. doi: 10.1128/IAI.00055-08
- Simser, J. A., Rahman, M. S., Dreher-Lesnick, S. M., and Azad, A. F. (2005). A novel and naturally occurring transposon, ISRpel in the *Rickettsia peacockii* genome disrupting the rickA gene involved in actin-based motility. *Mol. Microbiol.* 58, 71–79. doi: 10.1111/j.1365-2958.2005.04806.x
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039
- Ubeda, C., Tormo, M. A., Cucarella, C., Trotonda, P., Foster, T. J., Lasa, I., et al. (2003). Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol. Microbiol.* 49, 193–210. doi: 10.1046/j.1365-2958.2003.03577.x
- UniProt Consortium. (2015). UniProt: A hub for protein information. *Nucleic acids res.* 43, D204–D212. doi: 10.1093/nar/gku989
- Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369
- Wang, X., Ge, J., Liu, B., Hu, Y., and Yang, M. (2013). Structures of SdrD from *Staphylococcus aureus* reveal the molecular mechanism of how the cell surface receptors recognize their ligands. *Protein Cell* 4, 277–285. doi: 10.1007/s13238-013-3009-x
- Wästfelt, M., Stålhammar-Carlemalm, M., Delisse, A. M., Cabezon, T., and Lindahl, G. (1996). Identification of a family of *Streptococcal* surface proteins with extremely repetitive structure. *J. Biol. Chem.* 271, 18892–18897. doi: 10.1074/jbc.271.31.18892
- Webb, B., and Sali, A. (2014). Protein structure modeling with MODELLER. *Methods Mol. Biol.* 1137, 1–15. doi: 10.1007/978-1-4939-0366-5_1
- Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi: 10.1093/nar/gkm290
- Xiang, H., Feng, Y., Wang, J., Liu, B., Chen, Y., Liu, L., et al. (2012). Crystal structures reveal the multi-ligand binding mechanism of *Staphylococcus aureus* ClfB. *PLoS Pathog.* 8:e1002751. doi: 10.1371/journal.ppat.1002751
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Ziebuhr, W., Krimmer, V., Rachid, S., Lössner, I., Götz, F., and Hacker, J. (1999). A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: Evidence for control of the polysaccharide intercellular adhesion synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol. Microbiol.* 32, 345–356. doi: 10.1046/j.1365-2958.1999.01353.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Sun, Zhang, Zhou, Zhou, Li, Lin, Lu, Liu, Lu, Lin, Li, Xu, Zhu, Bao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.