

Comprehensive genomic landscape of antibiotic resistance in *Staphylococcus epidermidis*

Do-Hoon Lee,¹ Kihyun Lee,¹ Yong-Seok Kim,¹ Chang-Jun Cha¹

AUTHOR AFFILIATION See affiliation list on p. 11.

ABSTRACT *Staphylococcus epidermidis*, a common commensal bacterium found on human skin, can cause infections in clinical settings, and the presence of antibiotic resistance genes (ARGs) impedes the treatment of *S. epidermidis* infections. However, studies characterizing the ARGs in *S. epidermidis* with regard to genomic and ecological diversities are limited. Thus, we performed a comprehensive and comparative analysis of 405 high-quality *S. epidermidis* genomes, including those of 35 environmental isolates from the Han River, to investigate the genomic diversity of antibiotic resistance in this pathogen. Comparative genomic analysis revealed the prevalence of ARGs in *S. epidermidis* genomes associated with multi-locus sequence types. The genes encoding dihydrofolate reductase (*dfrC*) and multidrug efflux pump (*norA*) were genome-wide core ARGs. β -Lactam class ARGs were also highly prevalent in the *S. epidermidis* genomes, which was consistent with the resistance phenotype observed in river isolates. Furthermore, we identified chloramphenicol acetyltransferase genes (*cat*) in the plasmid-like sequences of the six river isolates, which have not been reported previously in *S. epidermidis* genomes. These genes were identical to those harbored by the *Enterococcus faecium* plasmids and associated with the insertion sequence 6 family transposases, homologous to those found in *Staphylococcus aureus* plasmids, suggesting the possibility of horizontal gene transfer between these Gram-positive pathogens. Comparison of the ARG and virulence factor profiles between *S. epidermidis* and *S. aureus* genomes revealed that these two species were clearly distinguished, suggesting genomic demarcation despite ecological overlap. Our findings provide a comprehensive understanding of the genomic diversity of antibiotic resistance in *S. epidermidis*.

IMPORTANCE A comprehensive understanding of the antibiotic resistance gene (ARG) profiles of the skin commensal bacterium and opportunistic pathogen *Staphylococcus epidermidis* needs to be documented from a genomic point of view. Our study encompasses a comparative analysis of entire *S. epidermidis* genomes from various habitats, including those of 35 environmental isolates from the Han River sequenced in this study. Our results shed light on the distribution and diversity of ARGs within different *S. epidermidis* multi-locus sequence types, providing valuable insights into the ecological and genetic factors associated with antibiotic resistance. A comparison between *S. epidermidis* and *Staphylococcus aureus* revealed marked differences in ARG and virulence factor profiles, despite their overlapping ecological niches.

KEYWORDS *Staphylococcus epidermidis*, antibiotic resistance, staphylococci, comparative genomics, virulence factor, *Staphylococcus aureus*

Staphylococci are classified based on their ability for blood plasma clotting for clinical diagnosis. *Staphylococcus aureus*, a classical pathogen responsible for a broad range of infections, is a typical coagulase-positive staphylococci (1). It can invade surgical and other wounds, sometimes causing sepsis. Despite its pathogenicity, *S. aureus* is usually

Editor Jack A. Gilbert, University of California San Diego, La Jolla, California, USA

Address correspondence to Chang-Jun Cha, cjcha@cau.ac.kr.

The authors declare no conflict of interest.

See the funding table on p. 12.

Received 15 February 2024

Accepted 4 April 2024

Published 10 May 2024

Copyright © 2024 Lee et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

harmless on the skin of approximately 30% of healthy individuals and survives well on inanimate surfaces, facilitating further infection (2).

Staphylococcus epidermidis, *S. haemolyticus*, and *S. saprophyticus* are well-known coagulase-negative staphylococci (CoNS). With the increase in the incidence of CoNS infection, there is growing awareness that they should be considered pathogens (3). CoNS members are more frequently isolated from patients in the intensive care unit than *S. aureus* during bloodstream infections (4). Historically, they have been considered culture contaminants, as they are widely distributed throughout the human body, including in the mouth and throat. These characteristics render the ability to distinguish between the infecting strains and normal bacterial flora difficult and recognition of their infection challenging. Most symptoms of CoNS infection differ markedly from those of *S. aureus* and are usually subacute or chronic with no apparent symptoms (5). CoNS infections rarely develop into life-threatening cases if treated promptly and appropriately, but in immunocompromised patients, fatal outcomes can occur (6).

S. epidermidis is a common human skin commensal bacterium isolated from the entire body surface of healthy individuals (7). Previously, *S. epidermidis* was considered a beneficial skin commensal bacterium involved in promoting homeostasis, defense against pathogens, and skin development (8). Currently, *S. epidermidis* is considered an opportunistic pathogen under certain conditions in which the skin barrier is disrupted by either genetic mutations or physical means (8, 9). The infections usually do not develop into life-threatening diseases due to their low toxicity compared to *S. aureus*; however, they are considered important in the clinical setting because of the high infection frequency and complexity of treatment required (9). Approximately 250,000 intravascular catheter-associated bloodstream infections are reported annually in the USA with *S. epidermidis* being the major cause of sepsis (10).

One of the challenges in treating infections caused by *S. epidermidis* is the presence of antibiotic resistance genes (ARGs). Clinical studies have revealed that antibiotic-resistant *S. epidermidis* strains associated with sepsis and orthopedic infections possess specific genotypes, and they are ubiquitous in clinical settings (11, 12). Notably, *S. epidermidis* strains isolated from patients with orthopedic infections harbor a higher proportion of ARGs and mobile genetic elements (MGEs) than strains isolated from healthy individuals or normal skin that are not the sites of infection (13). Furthermore, multi-locus sequence typing (MLST) analysis revealed the presence of 1,133 sequence types (STs) in *S. epidermidis*, indicating substantial strain-level diversity (14). However, previous studies on *S. epidermidis* have predominantly focused on clinical isolates, such as ST2 and ST23 strains of *S. epidermidis* associated with the nosocomial infection with limited representation of its genetic and ecological diversity. Considering the increasing prevalence of *S. epidermidis* infections in clinical settings, a genome-wide comparative analysis of *S. epidermidis* strains isolated from various sources should be performed to improve our understanding regarding the transmission dynamics of ARGs with the One-Health perspective (15, 16).

Here, we aimed to perform a comparative analysis of ARGs in 405 high-quality *S. epidermidis* genome sequences, including those from 35 isolates from the Han River, South Korea. Our study included both clinical and environmental isolates, enabling a more comprehensive evolutionary assessment of the genomic diversity of antibiotic resistance in *S. epidermidis*. We also compared the ARGs and virulence factors (VFs) present in *S. epidermidis* genomes with those in *S. aureus*, the results of which suggested genomic demarcation despite their ecological overlap.

RESULTS

Whole-genome sequences of *S. epidermidis* strains isolated from the Han River

In total, 35 strains of *S. epidermidis* were isolated from the Han River, South Korea. The 16S rRNA gene sequences of the isolates showed the highest similarities to *S. epidermidis* NCTC 11047^T (99.1%–100%). The whole-genome sequences were determined using the

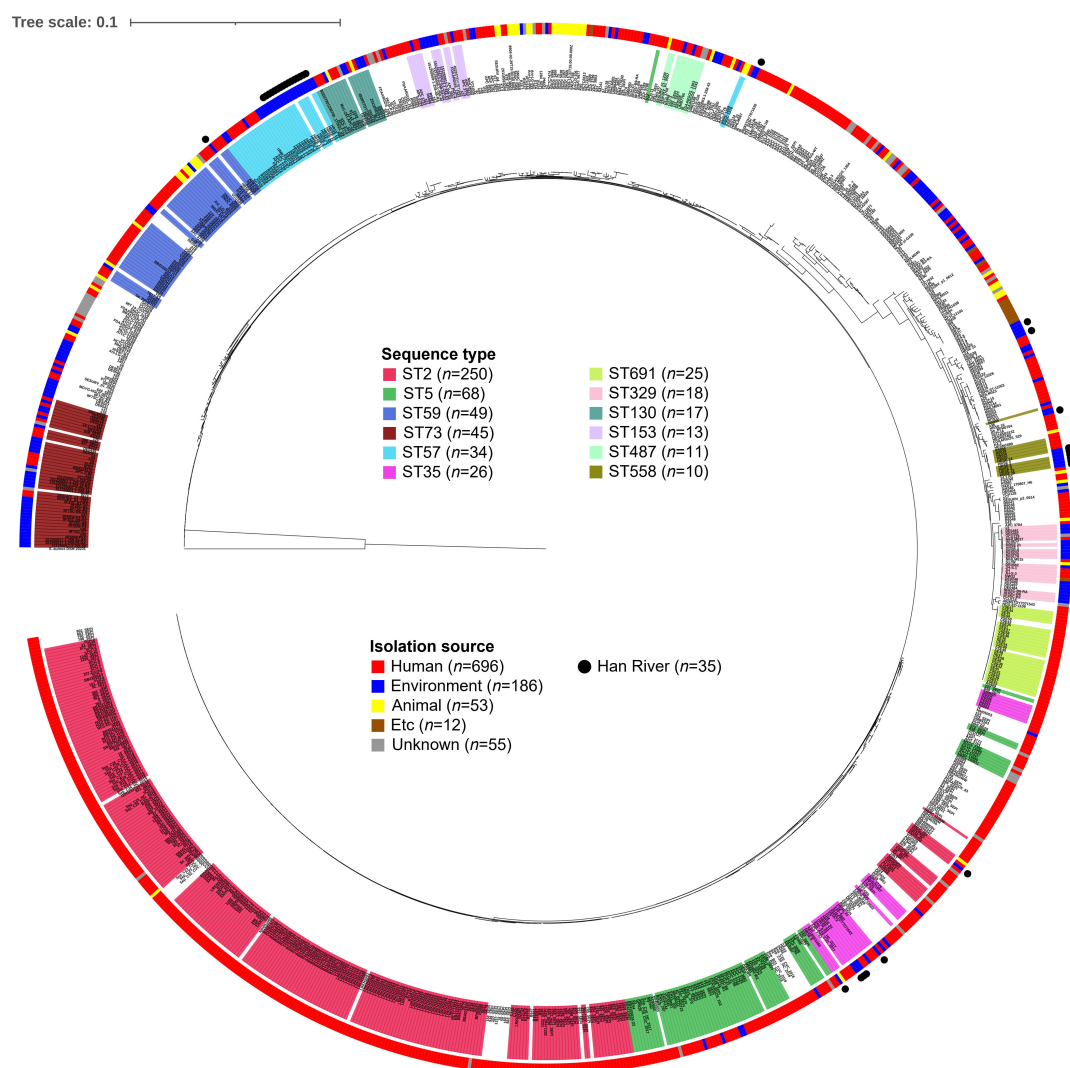


FIG 1 Whole-genome phylogeny of *Staphylococcus epidermidis* associated with their sequence types and isolation sources. A genome-based phylogenetic tree was reconstructed by the maximum-likelihood method using the concatenated alignments of 640 core genes among *S. epidermidis* genomes. *S. aureus* DSM 20231 was used as an outgroup. sequence types with $n \geq 10$ among the *S. epidermidis* genomes are color-shaded on strain names. Colors in the outer circle indicate the isolation source of each strain. Etc indicates isolation sources other than humans, animals, and the environment. Han River isolates are marked with black circles.

Oxford Nanopore Technologies sequencing platform. The assembled genomes were generally of high quality with a median of four contigs. The average nucleotide identity (ANI) values of the genomes with *S. epidermidis* NCTC 11047^T ranged from 99.5% to 96.8%, indicating that all strains were taxonomically identified to be *S. epidermidis* (17). The genomic features of *S. epidermidis* isolates are summarized in Table S1. The 35 draft genomes were approximately 2.5 ± 0.18 Mbp in size with an average GC content of 32.1%. Each genome contained $2,554 \pm 314$ genes, consisting of $2,474 \pm 314$ protein-coding genes, 19 rRNA genes, and 60 ± 1 tRNA genes (Table S1).

Phylogenomic analyses of *S. epidermidis*

A phylogenomic tree was constructed using the concatenated sequences of 640 core genes shared among the 1,002 *S. epidermidis* genomes (Fig. 1). The results showed that genome phylogeny was closely associated with the sequence type. Most *S. epidermidis* strains were isolated from humans, and ST2 was the most abundant lineage, which is known to spread globally in recent years (14). In contrast, hospital-acquired ST23 lineage,

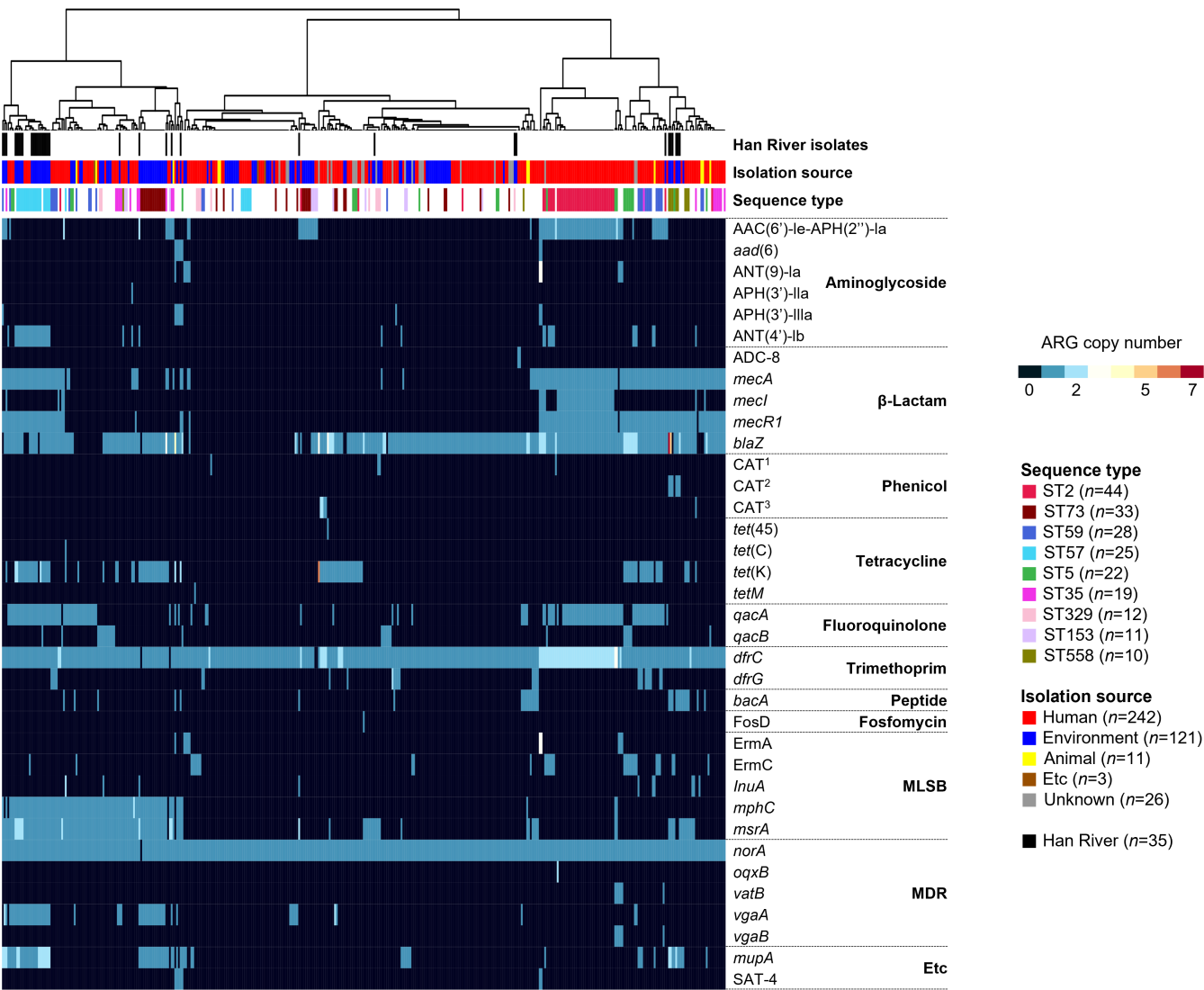


FIG 2 ARG profile of *Staphylococcus epidermidis* pangenome. The ARG profile-based clustering was performed using the Manhattan clustering method. Black color in the upper bar indicates the Han River isolates. Colors in the middle bar indicate the isolation sources of each strain. Etc indicates isolation sources other than humans, animals, and the environment. Colors in the lower bar indicate sequence types. CAT¹, CAT², and CAT³ indicate chloramphenicol acetyltransferase (CAT) from *Enterococcus faecalis*, *E. faecium*, and *S. intermedius*, respectively. The heatmap shows the copy number of ARGs.

which has recently emerged as a significant global threat to patient health, accounted for a very low percentage (0.7%) (12). The phylogeny of the Han River isolates showed wide distribution, belonging to ST57 ($n = 14$), ST35 ($n = 5$), ST558 ($n = 5$), ST2 ($n = 1$), ST59 ($n = 1$), ST329 ($n = 1$), ST1094 ($n = 1$), and ST1099 ($n = 1$), while six strains could not be classified using the MLST scheme.

ARG profiles of *S. epidermidis* genomes

The ARG profiles were determined for 405 *S. epidermidis* genomes with <50 contigs, and ARGs were detected in 403 genomes (Fig. 2). The genome-wide core ARGs, which were detected in more than 95% of *S. epidermidis* genomes, were identified to be the multidrug efflux pump-coding gene, *norA* (99% of 405 *S. epidermidis* genomes), and the trimethoprim resistance-coding gene, *dfrC* (98%) (Fig. 2). The genomes of *S. epidermidis* were found to harbor several different ARGs, including *blaZ* (76%), *mecA* (39%), *mecR1* (35%), and *mecI* (10%) genes conferring resistance to β-lactam antibiotics, and *msrA* (33%) and *mphC* (25%) genes conferring resistance to

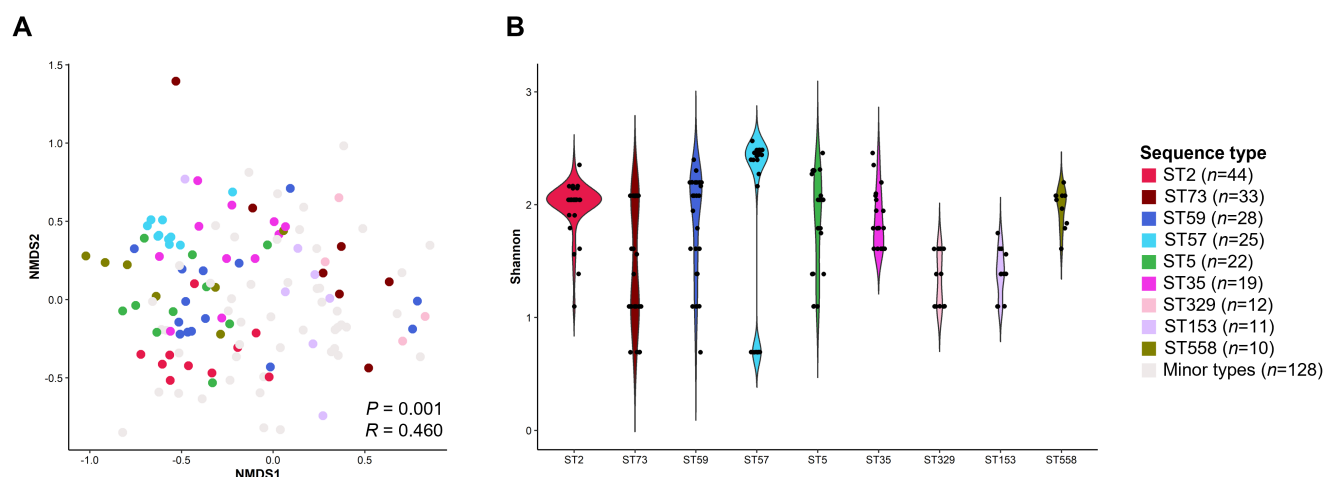


FIG 3 ARG diversity of *Staphylococcus epidermidis* associated with their sequence types. (A) Plot coordinates were determined by non-metric multidimensional scaling analysis of Bray–Curtis dissimilarity matrix, based on the ARG profiles. (B) The Shannon index was used to estimate the diversity of ARGs in each *S. epidermidis* genome. Minor types indicate the sum of each minor type ($1 < n < 10$).

macrolide-lincosamide-streptogramin B (MLSB) antibiotics. The *S. epidermidis* genomes also contained the fluoroquinolone resistance gene, *qacA* (31%), tetracycline resistance gene, *tet(K)* (24%), aminoglycoside resistance gene, *aac(6)-le-aph(2)-la* (20%), mupirocin resistance gene, *mupA* (18%), and the multidrug resistance gene, *vgaA* (14%).

To analyze the relationship between ARG profiles and sequence types (18), we selected genomes from sequence types with $n \geq 2$ among 403 *S. epidermidis* genomes that contained any of the ARGs. Distance-based non-metric multidimensional scaling (NMDS) analysis revealed that resistome clusters were significantly separated according to sequence type [analysis of similarities (ANOSIM), $P = 0.001$, $R = 0.460$] (Fig. 3A). The number of ARGs encoded per genome was distinct among the sequence types (Fig. 3B). Among the major sequence types, hospital-associated ST2 ($n = 44$) showed the most conserved ARG profiles, whereas ST73, ST59, ST57, and ST5 exhibited relatively diverse ARG profiles across genomes (Fig. 3B). The MGE profiles were also distinct among sequence types (for plasmid-like sequences, ANOSIM, $P = 0.001$, $R = 0.209$; for transposases, ANOSIM, $P = 0.001$, $R = 0.487$) (Fig. S1). Integrase sequences were not identified in the 403 *S. epidermidis* genomes.

The ARG profiles of the Han River isolates were similar to those of the entire *S. epidermidis* genomes obtained from the public database with a notable prevalence of β -lactam and MLSB class ARGs in most of the genomes (Fig. 4). The presence of ARGs against β -lactams, phenicol, tetracycline, and MLSB was highly consistent with the resistance phenotypes of these isolates, whereas that of ARGs against aminoglycosides, fluoroquinolones, and trimethoprim was not (Fig. 4). Notably, the plasmid-like sequences of the Han River isolates belonging to ST558 ($n = 5$) and ST1099 ($n = 1$) harbored chloramphenicol acetyltransferase genes (*cat*) identical to those found in the plasmids of *Enterococcus faecium* (Fig. 5). These plasmid-like sequences were associated with insertion sequence 6 family transposases, homologous to those found in the plasmids of *S. aureus* (Fig. 5), which do not contain *cat* gene.

Comparison of MLST, ARG, MGE, and VF profiles among *S. epidermidis* genomes of human, animal, and environmental isolates

MLST distribution was compared among human, animal, and environmental isolates of *S. epidermidis*. As a result, MLST composition significantly differed according to isolation sources (Fisher's exact test, $P < 0.001$; Cramer's $V = 0.492$) (Fig. 6A): ST2, ST59, and ST5 were significantly more abundant in *S. epidermidis* genomes of human origin than

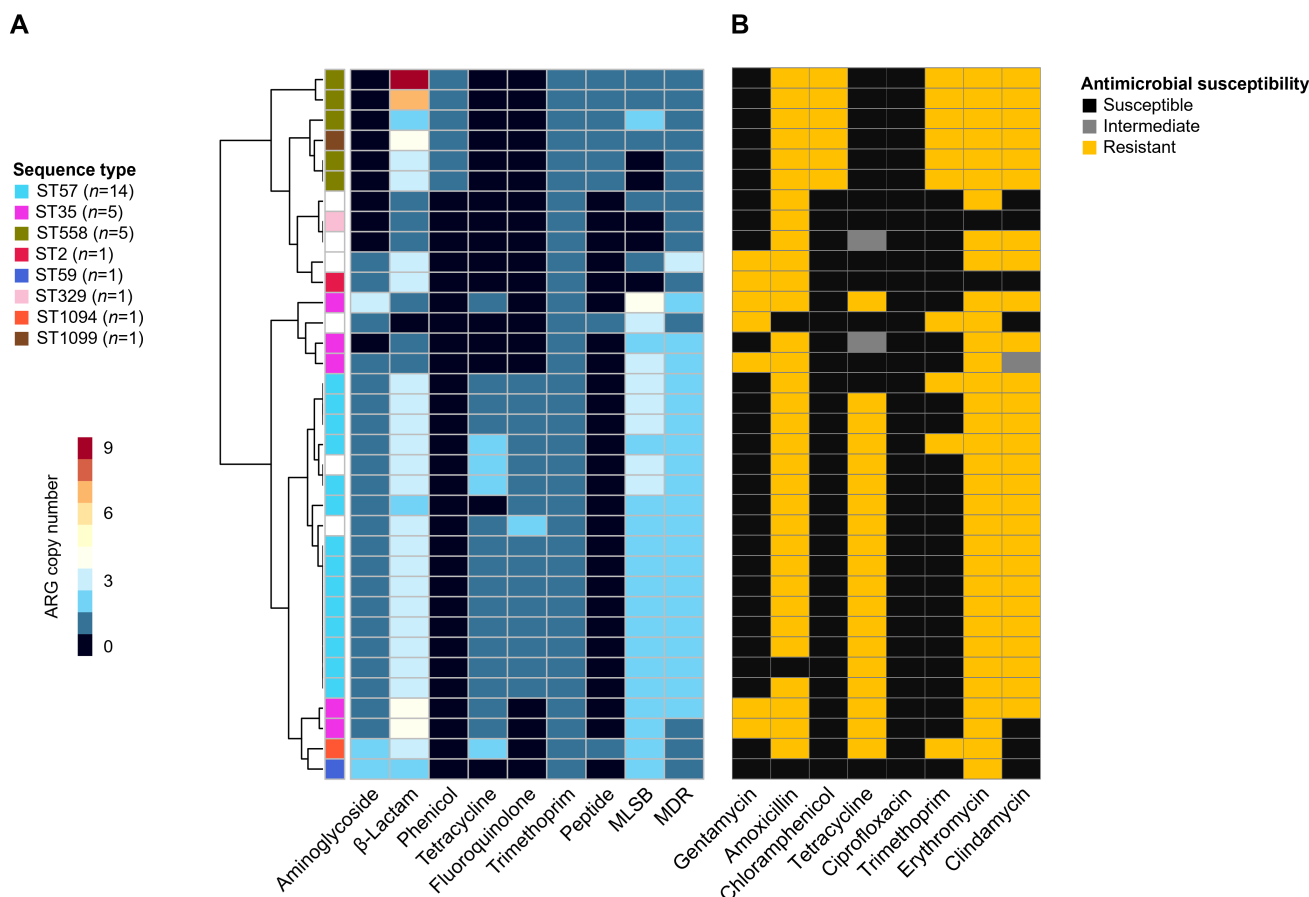


FIG 4 Genome-wide ARG profiles and resistance phenotypes of *Staphylococcus epidermidis* from the Han River isolates. (A) ARG profiles of *S. epidermidis* genomes from Han River isolates. The ARG profile-based clustering was performed using the Manhattan clustering method. Colors on the right side of the tree nodes indicate sequence types. The heatmap shows the copy number of the ARGs. (B) Antimicrobial susceptibility test results for the Han River isolates. Black, gray, and yellow colors indicate susceptible, intermediate, and resistant, respectively.

environmental genomes, whereas ST73 was highly dominant in environmental genomes. The ARG profiles of *S. epidermidis* genomes were clustered, but not clearly separated, among humans, animals, and the environment (ANOSIM, $P = 0.001$, $R = 0.109$) (Fig. 6B). Clusters of animals and environment were included within the human clusters, suggesting that most *S. epidermidis* strains were of human origin. Notably, ARG profiles of *S. epidermidis* genomes were distinguished among six different isolation countries ($n \geq 24$) (ANOSIM, $P = 0.001$, $R = 0.232$) (Fig. S2). In the case of MGEs, the transposase profiles were clustered among different isolation sources (ANOSIM, $P = 0.002$, $R = 0.055$) (Fig. S3A), whereas the plasmid-like sequences were not significantly clustered (ANOSIM, $P = 0.362$, $R = 0.001$) (Fig. S3B). In addition, the VF profiles of *S. epidermidis* genomes were not associated with isolation sources (ANOSIM, $P = 0.315$, $R = 0.009$) (Fig. S3C).

Comparison of the ARG and VF profiles between *S. epidermidis* and *S. aureus* genomes

Phylogenetically adjacent *S. epidermidis* and *S. aureus* are both common commensals on human skin and are major causes of nosocomial infections. The overlapping ecological niches of these two species suggest that various inter-species interactions may occur. Therefore, we performed a comparative genomic analysis focusing on ARGs and VFs. Our results revealed that each species has a distinct ARG profile with certain genes found exclusively in each genome (Fig. 7A and B). *S. epidermidis* genomes displayed

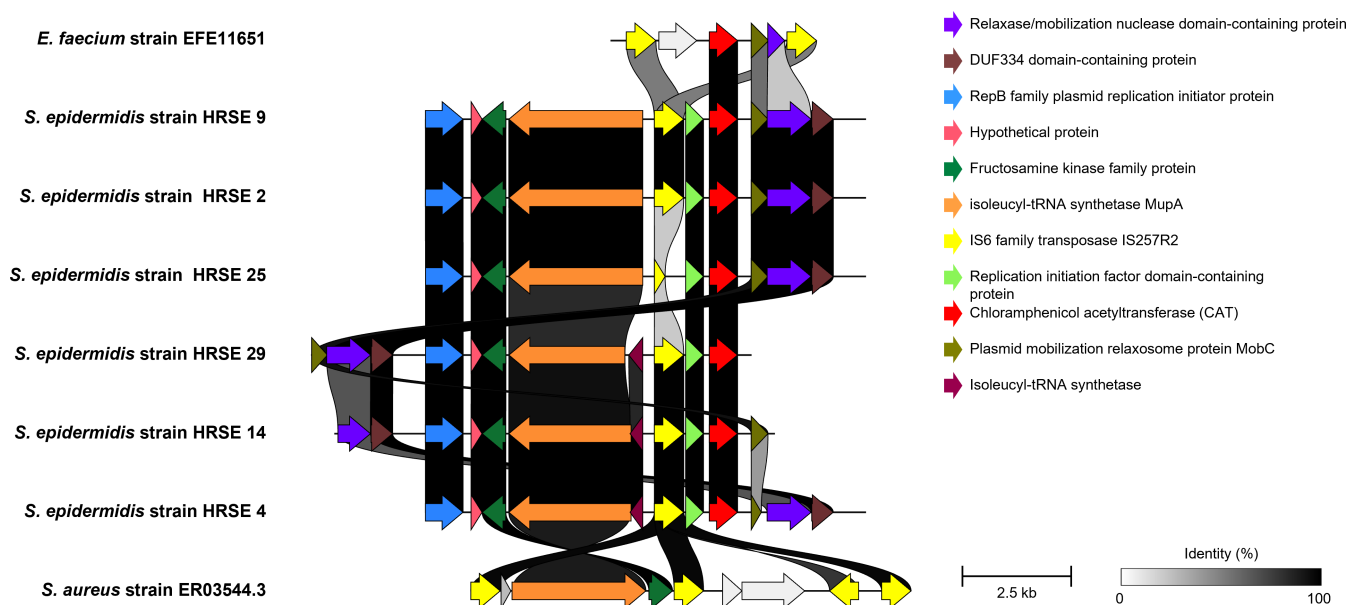


FIG 5 Plasmid-like sequences of the Han River isolates harboring chloramphenicol acetyltransferase genes (*cat*) and associated with IS6 family transposases. Pairwise alignment and visualization of gene clusters were performed using Clinker.

relatively more diverse ARG profiles than *S. aureus* genomes (Fig. 7C). *S. aureus* possessed a larger number of ARGs per genome (9.4 ARG copies) than *S. epidermidis* (6.2 ARG copies) (Fig. 7D). The genome-wide core ARGs in *S. epidermidis* were *dfrC* and *norA*, while *lmrS*, *S. aureus* *norA*, *mepA*, and *tet(38)* were the core ARGs in the *S. aureus* genomes; core ARGs were not shared between the two species (Fig. 7A). The *S. epidermidis* and *S. aureus* genomes encoded the β -lactam class ARGs, *blaZ* (76% and 75% of each genome, respectively), *mecA* (39% and 60%), *mecR1* (35% and 53%), and *mecI* (10% and 23%), and the MLSB class ARGs, *msrA* (33% and 16%) and *mphC* (25% and 15%) (Fig. 7A).

The VF profiles of *S. epidermidis* and *S. aureus* were more clearly distinguishable than their ARG profiles (Fig. S4A and B). The genome-wide core VFs of *S. epidermidis* included *tufa* (100% of *S. epidermidis* genomes), *clpP* (99%), *acpXL* (99%), and *aur* (97%). In contrast, *S. aureus* consisted of 53 core VFs, including the four core VFs of *S. epidermidis* (Fig. S4A), indicating that the VF composition of *S. aureus* genomes was considerably more diverse (Fig. S4C) and abundant than that of *S. epidermidis* genomes. Notably, *esaG*, which is associated with the type VII secretion system, was the most abundant VF across all genomes (99% of *S. aureus* genomes) (Fig. S4A). *S. aureus* possessed an enormously larger number of VFs per genome (80.8 VF copies) than *S. epidermidis* (4.8 VF copies) (Fig. S4D).

DISCUSSION

The increasing prevalence of multidrug-resistant *S. epidermidis* has become a serious problem constraining treatment options in clinical settings. It has a relatively small genome of approximately 2.5 Mb, and 20% of its genome is composed of variable structures (19), which suggests that this opportunistic pathogen has an open pan-genome and high potential to acquire new genetic traits via MGEs (19). Furthermore, evidence of horizontal gene transfer between *S. epidermidis* genomes from different skin sites involving antibiotic resistance and virulence genes has been reported (20). The flexible genome structures of *S. epidermidis* and horizontal gene transfer contribute significantly to the unpredictability of the *S. epidermidis* genome (7).

In this study, we highlighted the ARGs present in *S. epidermidis* genomes by analyzing comprehensive and high-quality genomes obtained from diverse habitats. Previous genomic studies have focused only on the clinical isolates, which may lead to biased

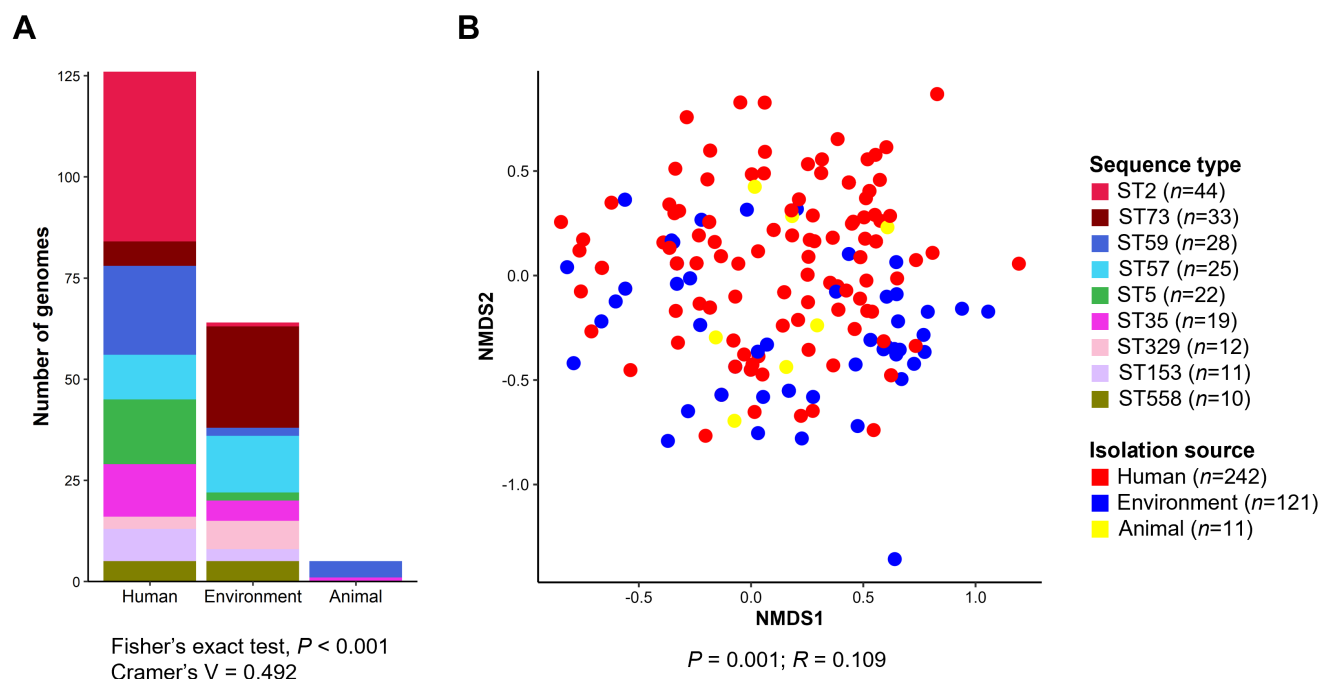


FIG 6 MLST distribution and ARG profiles of *Staphylococcus epidermidis* genomes among isolation sources. (A) MLST distribution among different isolation sources was analyzed using Fisher's exact test and Cramer's V. (B) Plot coordinates were determined by non-metric multidimensional scaling analysis of Bray–Curtis dissimilarity matrix, based on the ARG profiles.

views and hinder our understanding of the genetic structures of antibiotic resistance across different habitats. Therefore, a comparison of the entire *S. epidermidis* genomes from the One-Health perspective is critical. Although *S. epidermidis* is primarily found on human skin, 33% of the genomes analyzed in our study was obtained from various habitats other than human skin, including the environment (30%) and animals (3%). In the present study, we isolated 35 environmental strains of *S. epidermidis* from the Han River and those genomes were analyzed. The ARG profiles of these isolates were similar to those of the entire *S. epidermidis* genomes obtained from the public database. The genome-predicted ARG profiles were not perfectly consistent with experimentally determined resistance phenotype (Fig. 4). This inconsistency may be due to several reasons: ARGs may not be expressed under certain conditions. ARGs may be mutated and become non-functional. The presence of multidrug efflux pumps can also make resistance phenotype to certain antibiotics unpredictable. ARGs not yet discovered may confer resistance phenotype. Interestingly, genomes of the six environmental isolates from the Han River sequenced in this study contained plasmid-like sequences harboring *cat* genes identical to those of *E. faecium*. These ARGs have never been identified in *S. epidermidis* genomes. Furthermore, these sequences were associated with IS6 family transposases, which are homologous to those of *S. aureus*. These results suggest that ARGs structured in the transposon of plasmid may have been transmitted between these Gram-positive pathogens (21, 22).

Previous studies have suggested that the genotype of *S. epidermidis* is associated with infection and host disease (12, 23), and its ARG profile may also be closely related to its sequence type (12, 24). Our results revealed that *S. epidermidis* ARG profiles based on their genomes could be distinguished according to different sequence types. Hospital-associated ST2 was found to have relatively conserved ARG profiles, whereas sequence types isolated from various habitats, such as ST73, ST59, ST57, and ST5, possessed more diverse ARG profiles, suggesting that ARG diversity within a specific sequence type may be influenced by ecological niches.

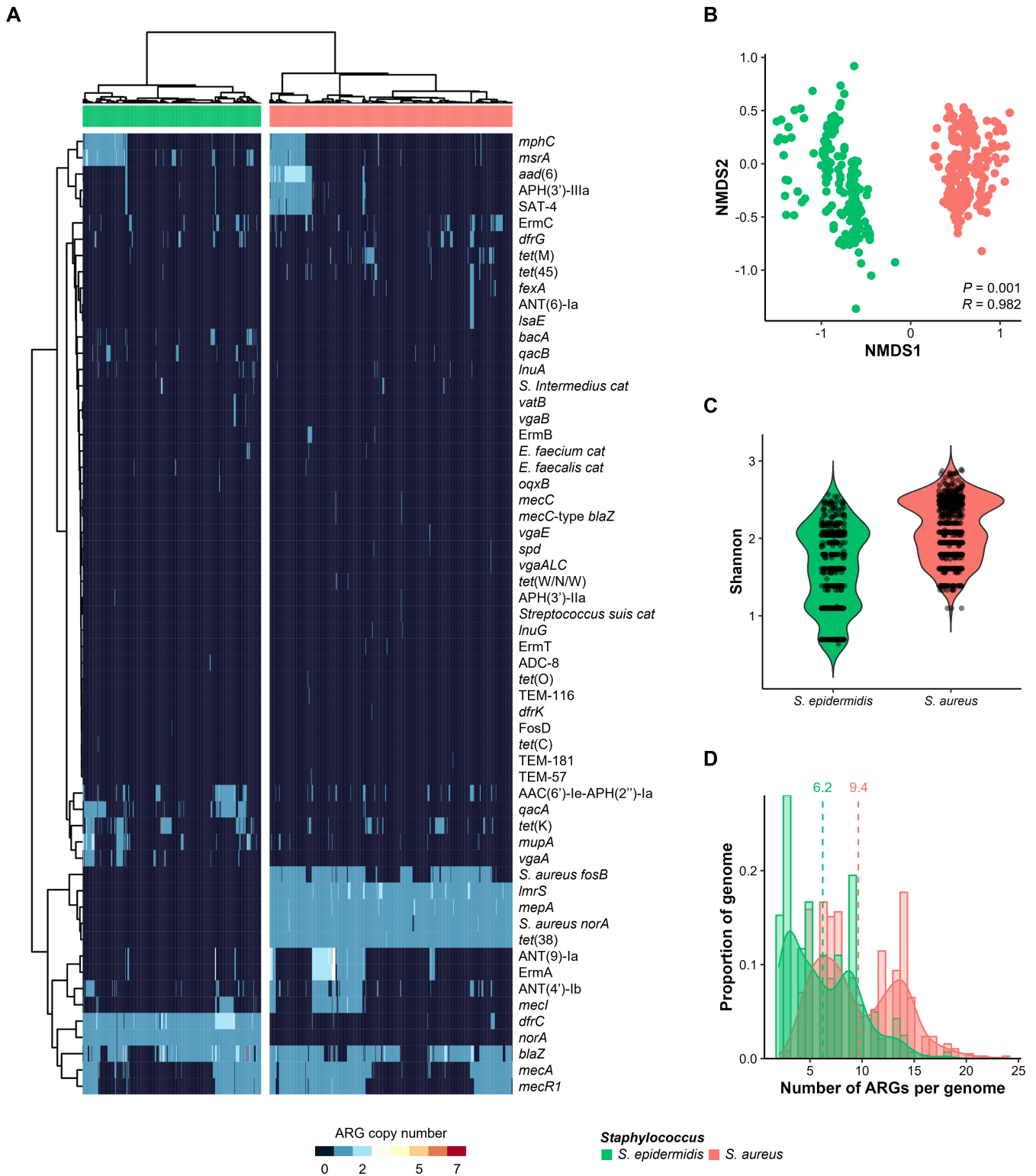


FIG 7 Comparison of ARG profiles between *Staphylococcus epidermidis* and *S. aureus* genomes. (A) The ARG profile-based clustering was performed using the Manhattan clustering method. The heatmap shows the copy numbers of ARGs. (B) The coordinates of the plot were determined by a non-metric multidimensional scaling analysis of Bray–Curtis dissimilarity matrix, based on ARG profiles. (C) The ARG diversity in each genome was estimated using the Shannon index. (D) Distribution of ARG copies per genome. The dotted lines indicate the average number of ARG copies in each species. Green and coral colors indicate *S. epidermidis* and *S. aureus*, respectively.

Our study revealed that the ARG and VF profiles of *S. epidermidis* and *S. aureus* could be clearly distinguished at the genome level. Despite overlapping ecological niches, gene transfer between staphylococcal species is believed to be rare owing to the presence of restriction modification systems and clustered regularly interspaced short palindromic repeats (CRISPR) loci that degrade foreign DNAs (25, 26). However, several studies have indicated that the proportion of CRISPR loci within staphylococci is lower than that reported previously (27). In addition, multiple lines of evidence regarding the inter-species transfer of genes related to antibiotic resistance, pathogenicity, and metal resistance between *S. epidermidis* and *S. aureus* have been documented (22, 28, 29). The transfer of genes may herald the emergence of *S. aureus* lineages with increased antibiotic resistance and virulence, as the genomic flexibility of *S. epidermidis* renders it likely to act as a genetic reservoir for pathogens such as *S. aureus* (20, 30).

In conclusion, our results underscore that the genome-wide ARG profiles of *S. epidermidis* are well characterized among their sequence types. Our study contributes to a more comprehensive understanding of *S. epidermidis* antibiotic resistance at the pangenome level and provides a genomic basis for variations in resistance and virulence between staphylococcal pathogens *S. epidermidis* and *S. aureus*. Addressing how and why these two species from a similar ecological niche have evolved differently to possess distinct antibiotic resistance and virulence genes are the next steps for elucidating genomic demarcation.

MATERIALS AND METHODS

Isolation and identification of *S. epidermidis*

Twenty-two *S. epidermidis* strains were isolated from the Han River, South Korea, from 2016 to 2019 using Mueller–Hinton agar (Difco, Detroit, MI, USA) supplemented with various antibiotics. The antibiotics used were amoxicillin (16 mg/L), chloramphenicol (16 mg/L), erythromycin (2 mg/L), gentamicin (8 mg/L), lincomycin (2 mg/L), sulfamethoxazole (40 mg/L), and tetracycline (4 mg/L). All antibiotics were obtained from Sigma-Aldrich (St. Louis, MO, USA). Thirteen strains were isolated from the Han River in May 2019 using the Baird–Parker medium supplemented with egg yolk (Difco). These isolates were identified using the EzBioCloud database (https://www.ezbiocloud.net/resources/16s_download) (31) based on their 16S rRNA gene sequences.

Antimicrobial susceptibility test

Antimicrobial susceptibility test was performed using the disk diffusion assay according to the EUCAST guidelines, 2016. The amount of antibiotic in each disk was as follows: amoxicillin (10 µg), cephalexin (30 µg), chloramphenicol (30 µg), ciprofloxacin (5 µg), clindamycin (2 µg), colistin (10 µg), erythromycin (15 µg), fosfomycin (200 µg), gentamicin (10 µg), linezolid (10 µg), meropenem (10 µg), streptomycin (10 µg), sulfamethoxazole (50 µg), rifampicin (5 µg), tetracycline (30 µg), trimethoprim (5 µg), tylosin (30 µg), and vancomycin (30 µg). All disks were obtained from Liofilchem (Roseto, Italy).

Whole-genome sequencing and assembly

A total of 35 environmental isolates from the Han River were subjected to whole-genome sequencing. Genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) and prepared using the ligation sequencing kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK), following the manufacturer's protocols. Whole-genome sequencing was performed using Oxford Nanopore MinION flow cells (R9.4.1 FLO-MIN106; Oxford Nanopore Technologies) at DNA Link (Seoul, Korea). Basecalling on the raw signal data was performed using the Guppy v6.0.7. The draft genomes were *de novo* assembled using Flye, Miniasm, and Raven implemented in Tricycler v0.5.3 (32).

Phylogenomic analysis

The genome assembly data of 967 *S. epidermidis* strains and *S. aureus* DSM 20231 (an outgroup) obtained from the NCBI RefSeq database, and 35 *S. epidermidis* genomes from the environmental isolates sequenced in this study, were used for phylogenomic analysis. A phylogenomic tree was constructed using the core genes shared among all compared genomes. To identify the core genes, a bidirectional best BLASTP search was performed for protein-coding sequences across the entire genome. A gene was considered to be a core gene when the best hits for that gene were found in more than 95% of the compared genomes. The nucleotide sequences of each core gene were aligned using MAFFT v7 (33), and the resulting alignments were concatenated. Maximum-likelihood phylogenetic trees were constructed from the concatenated genes by running FastTree2 as a general-time reversible model of nucleotide substitutions (34). Interactive Tree of Life v5 was used to visualize the phylogenomic tree (35). MLST was performed using the MLST software (36) based on the PubMLST database. Accurate taxonomic identification was performed using ANI values obtained using the FastANI tool (37).

Analysis of ARGs, VFs, and MGEs in the genome sequences

For the analysis of ARGs, VFs, and MGEs, 405 genome sequences (Data Set S1) were used, including 370 *S. epidermidis* genomes with <50 contigs selected from the NCBI RefSeq database (Data Set S2) and 35 river genomes sequenced in this study. For *S. aureus*, 550 genomes were randomly selected. ARGs were detected based on the Comprehensive Antibiotic Resistance Database using the Resistance Gene Identifier software with strict cutoff values (38). Shannon diversity indices of the ARG and VF profiles were estimated using the vegan R package (39). Clinker was used to visualize gene clusters (40). VFs were annotated using a BLASTP search with the core data set of the virulence factor database with a cutoff of $\geq 70\%$ identity and $\geq 70\%$ query coverage. The ARG and VF profiles were clustered using a segmentation algorithm that clustered points based on the Manhattan distance in the pheatmap R package (41). A heatmap was constructed using the pheatmap R package (41). Transposases were annotated using a BLASTP search and the IS database with a cutoff of $\geq 80\%$ identity and $\geq 80\%$ query coverage. Plasmid-like sequences were identified using a BLASTN search against plasmid sequences downloaded from the NCBI RefSeq database with $\geq 80\%$ identity and $\geq 2,000$ bp alignment length. Differences in ARG, VF, and MGE profiles among each sequence type, isolation source, isolation country, or species were analyzed using NMDS, and significance was tested using the ANOSIM in the vegan R package (39). MLST distribution among different isolation sources was analyzed using Fisher's exact test in R and Cramer's V in the rcompanion R package (42, 43).

ACKNOWLEDGMENTS

C.-J.C. designed the study. D.-H.L., K.L., and Y.-S.K. performed experiments. D.-H.L. and K.L. analyzed data. D.-H.L. and C.-J.C. wrote the manuscript.

This study was supported by the Korea Ministry of Environment (MOE) as The Environmental Health Action Program (2016001350004) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2023R1A2C1003654).

We thank J. Kim for technical assistance at the BT Research Facility Center, Chung-Ang University.

AUTHOR AFFILIATION

¹Department of Systems Biotechnology and Center for Antibiotic Resistome, Chung-Ang University, Anseong, South Korea

PRESENT ADDRESS

Kihyun Lee, CJ Bioscience, Seoul, South Korea

AUTHOR ORCID*s*

Do-Hoon Lee  <http://orcid.org/0000-0001-5816-3097>

Chang-Jun Cha  <http://orcid.org/0000-0002-2210-2898>

FUNDING

Funder	Grant(s)	Author(s)
Ministry of Environment (MOE)	2016001350004	Chang-Jun Cha
National Research Foundation of Korea (NRF)	NRF-2023R1A2C1003654	Chang-Jun Cha

AUTHOR CONTRIBUTIONS

Do-Hoon Lee, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Kihyun Lee, Formal analysis, Investigation | Yong-Seok Kim, Investigation | Chang-Jun Cha, Conceptualization, Data curation, Funding acquisition, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Raw nucleotide sequence reads and assembled genome sequences of the Han River isolates were deposited in the NCBI BioProject under the accession number [PRJNA998543](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA998543).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Data Set S1 (mSystems00226-24-S0001.xlsx). *Staphylococcus epidermidis* genome assemblies used for ARG, VF, and MGE analysis.

Data Set S2 (mSystems00226-24-S0002.xlsx). Detailed information of 370 genomes downloaded from RefSeq.

Supplemental figures and table (mSystems00226-24-S0003.pdf). Fig. S1 to S4, Table S1.

REFERENCES

1. Lowy FD. 1998. *Staphylococcus aureus* infections. N Engl J Med 339:520–532. <https://doi.org/10.1056/NEJM199808203390806>
2. Solberg CO. 1965. A study of carriers of *Staphylococcus aureus* with special regard to quantitative bacterial estimations. Acta Med Scand Suppl 436:1–96.
3. Diekema DJ, Pfaller MA, Schmitz FJ, Smayevsky J, Bell J, Jones RN, Beach M, SENTRY Participants Group. 2001. Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY antimicrobial surveillance program, 1997–1999. Clin Infect Dis 32 Suppl 2:S114–32. <https://doi.org/10.1086/320184>
4. System N. 1999. Data summary from January 1990–may 1999, issued June 1999. a report from the NNIS system. Am J Infect Control 27:520–532. [https://doi.org/10.1016/S0196-6553\(99\)70031-3](https://doi.org/10.1016/S0196-6553(99)70031-3)
5. von Eiff C, Peters G, Heilmann C. 2002. Pathogenesis of infections due to coagulase-negative *Staphylococci*. The Lancet Infectious Diseases 2:677–685. [https://doi.org/10.1016/S1473-3099\(02\)00438-3](https://doi.org/10.1016/S1473-3099(02)00438-3)
6. Sabaté Brescó M, Harris LG, Thompson K, Stanic B, Morgenstern M, O'Mahony L, Richards RG, Moriarty TF. 2017. Pathogenic mechanisms and host interactions in *Staphylococcus epidermidis* device-related infection. Front Microbiol 8:1401. <https://doi.org/10.3389/fmicb.2017.01401>
7. Severn MM, Horswill AR. 2023. *Staphylococcus epidermidis* and its dual lifestyle in skin health and infection. Nat Rev Microbiol 21:97–111. <https://doi.org/10.1038/s41579-022-00780-3>
8. Brown MM, Horswill AR. 2020. *Staphylococcus epidermidis*–skin friend or foe PLoS Pathog 16:e1009026. <https://doi.org/10.1371/journal.ppat.1009026>
9. Vuong C, Otto M. 2002. *Staphylococcus epidermidis* infections. Microbes Infect 4:481–489. [https://doi.org/10.1016/s1286-4579\(02\)01563-0](https://doi.org/10.1016/s1286-4579(02)01563-0)
10. Rogers KL, Fey PD, Rupp ME. 2009. Coagulase-negative staphylococcal infections. Infect Dis Clin North Am 23:73–98. <https://doi.org/10.1016/j.idc.2008.10.001>
11. Krediet TG, Mascini EM, van Rooij E, Vlooswijk J, Paaauw A, Gerards LJ, Fleer A. 2004. Molecular epidemiology of coagulase-negative

- staphylococci causing sepsis in a neonatal intensive care unit over an 11-year period. *J Clin Microbiol* 42:992–995. <https://doi.org/10.1128/JCM.42.3.992-995.2004>
12. Lee JYH, Monk IR, Gonçalves da Silva A, Seemann T, Chua KYL, Kearns A, Hill R, Woodford N, Bartels MD, Strommenger B, Laurent F, Dodémont M, Deplano A, Patel R, Larsen AR, Korman TM, Stinear TP, Howden BP. 2018. Global spread of three multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat Microbiol* 3:1175–1185. <https://doi.org/10.1038/s41564-018-0230-7>
 13. Månsson E, Bech Johannesen T, Nilsdotter-Augustinsson Å, Söderquist B, Stegger M. 2021. Comparative genomics of *Staphylococcus epidermidis* from prosthetic-joint infections and nares highlights genetic traits associated with antimicrobial resistance, not virulence. *Microb Genom* 7:000504. <https://doi.org/10.1099/mgen.0.000504>
 14. Miragaia M, Thomas JC, Couto I, Enright MC, de Lencastre H. 2007. Inferring a population structure for *Staphylococcus epidermidis* from Multilocus sequence typing data. *J Bacteriol* 189:2540–2552. <https://doi.org/10.1128/JB.01484-06>
 15. Kim DW, Cha CJ. 2021. Antibiotic resistome from the one-health perspective: understanding and controlling antimicrobial resistance transmission. *Exp Mol Med* 53:301–309. <https://doi.org/10.1038/s12276-021-00569-z>
 16. Djordjevic SP, Jarocki VM, Seemann T, Cummins ML, Watt AE, Drigo B, Wyrsh ER, Reid CJ, Donner E, Howden BP. 2024. Genomic surveillance for antimicrobial resistance - a one health perspective. *Nat Rev Genet* 25:142–157. <https://doi.org/10.1038/s41576-023-00649-y>
 17. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106:19126–19131. <https://doi.org/10.1073/pnas.0906412106>
 18. Lee K, Kim DW, Cha CJ. 2021. Overview of bioinformatic methods for analysis of antibiotic resistome from genome and metagenome data. *J Microbiol* 59:270–280. <https://doi.org/10.1007/s12275-021-0652-4>
 19. Conlan S, Mijares LA, Program NCS, Becker J, Blakesley RW, Bouffard GG, Brooks S, Coleman H, Gupta J, Gurson N, Park M, Schmidt B, Thomas PJ, Otto M, Kong HH, Murray PR, Segre JA. 2012. *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol* 13:R64. <https://doi.org/10.1186/gb-2012-13-7-r64>
 20. Zhou W, Spoto M, Hardy R, Guan C, Fleming E, Larson PJ, Brown JS, Oh J. 2020. Host-specific evolutionary and transmission dynamics shape the functional diversification of *Staphylococcus epidermidis* in human skin. *Cell* 180:454–470. <https://doi.org/10.1016/j.cell.2020.01.006>
 21. Zhu W, Clark N, Patel JB. 2013. Psk41-like plasmid is necessary for Inc18-like *vanA* Plasmid transfer from *Enterococcus faecalis* to *Staphylococcus aureus* in vitro. *Antimicrob Agents Chemother* 57:212–219. <https://doi.org/10.1128/AAC.01587-12>
 22. Méric G, Miragaia M, de Been M, Yahara K, Pascoe B, Mageiros L, Mikhail J, Harris LG, Wilkinson TS, Rolo J, Lambie S, Bray JE, Jolley KA, Hanage WP, Bowden R, Maiden MCJ, Mack D, de Lencastre H, Feil EJ, Corander J, Sheppard SK. 2015. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biol Evol* 7:1313–1328. <https://doi.org/10.1093/gbe/evv066>
 23. Méric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, Kittiwon N, Tadee P, Post V, Lambie S, Bowden R, Bray JE, Morgenstern M, Jolley KA, Maiden MCJ, Feil EJ, Didelot X, Miragaia M, de Lencastre H, Moriarty TF, Rohde H, Massey R, Mack D, Corander J, Sheppard SK. 2018. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat Commun* 9:5034. <https://doi.org/10.1038/s41467-018-07368-7>
 24. Miragaia M, Couto I, de Lencastre H. 2005. Genetic diversity among methicillin-resistant *Staphylococcus epidermidis* (MRSE). *Microb Drug Resist* 11:83–93. <https://doi.org/10.1089/mdr.2005.11.83>
 25. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–1845. <https://doi.org/10.1126/science.1165771>
 26. Lee JYH, Carter GP, Pidot SJ, Guérillot R, Seemann T, Gonçalves da Silva A, Foster TJ, Howden BP, Stinear TP, Monk IR. 2019. Mining the methylome reveals extensive diversity in *Staphylococcus epidermidis* restriction modification. *mBio* 10. <https://doi.org/10.1128/mBio.02451-19>
 27. Rossi CC, Souza-Silva T, Araújo-Alves AV, Giambiagi-deMarval M. 2017. CRISPR-CAS systems features and the gene-reservoir role of coagulase-negative staphylococci. *Front Microbiol* 8:1545. <https://doi.org/10.3389/fmicb.2017.01545>
 28. Ray MD, Boundy S, Archer GL. 2016. Transfer of the methicillin resistance genomic Island among staphylococci by conjugation. *Mol Microbiol* 100:675–685. <https://doi.org/10.1111/mmi.13340>
 29. Banaszkievicz S, Calland JK, Mourkas E, Sheppard SK, Pascoe B, Bania J. 2019. Genetic diversity of composite enterotoxigenic *Staphylococcus epidermidis* pathogenicity islands. *Genome Biol Evol* 11:3498–3509. <https://doi.org/10.1093/gbe/evz259>
 30. Otto M. 2013. Coagulase-negative staphylococci as reservoirs of genes facilitating MRSA infection. *BioEssays* 35:4–11. <https://doi.org/10.1002/bies.201200112>
 31. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing ezbiocloud: a taxonomically United database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1617. <https://doi.org/10.1099/ijsem.0.001755>
 32. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Tricycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 22:266. <https://doi.org/10.1186/s13059-021-02483-z>
 33. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
 34. Price MN, Dehal PS, Arkin AP. 2010. Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>
 35. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) V5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
 36. Seemann T. 2016. MLST: scan contig files against PubMLST typing schemes. <https://github.com/tseemann/mlst>
 37. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
 38. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>
 39. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2020. Vegan: community ecology package. Available from: <https://github.com/vegandevs/vegan>
 40. Gilchrist CLM, Chooi YH. 2021. Clinker & Clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37:2473–2475. <https://doi.org/10.1093/bioinformatics/btab007>
 41. Kolde R. 2012. Pheatmap: Pretty Heatmaps. v1.0.12: <https://CRAN.R-project.org/package=pheatmap>
 42. R Core Team R. 2013. R: a language and environment for statistical computing.
 43. Mangiafico S, Mangiafico MS. 2017. Package ‘rcompanion’. <https://CRAN.R-project.org/package=rcompanion>.