Comparative genomic and pan-genomic characterisation of

Staphylococcus epidermidis from different sources unveils the molecular

basis and potential biomarkers of pathogenic strains

Shudan Lin^{b*}, Bianjin Sun^{a*}, Xinrui Shi^{a*}, Yi Xu^a, Yunfeng Gu^a, Xiaobin Gu^b, Xueli Ma^a,

Tian Wan^a, Jie Xu^a, Jianzhong Su^a, Yongliang Lou^{b#}, Meiqin Zheng^{a,b#}

a.School of Ophthalmology and Optometry and Eye Hospital, School of Biomedical

Engineering, Wenzhou Medical University, Wenzhou, Zhejiang, China

b.Zhejiang Provincial Key Laboratory for Technology and Application of Model

Organisms, Key Laboratory of Laboratory Medicine, Ministry of Education, School of

Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou, Zhe

jiang, China

*These authors contributed equally to this work and should be considered co-first

authors.

#Co-Correspondence to:

Yongliang Lou, Professor, Tel: 86-577-86699200, E-mail: louyongliang2013@163.com;

Meiqin Zheng, Professor, TEL: 86-577-88068857, E-mail: zmqylh@126.com;

Abstract Coagulase-negative *Staphylococcus* (CoNS) is the most common pathogen causing traumatic endophthalmitis, *Staphylococcus epidermidis* is the most common species which colonizes human skin, eye surfaces and nasal cavity and is the main cause of nosocomial infection, specially foreign body-related bloodstream infections (FBR-BSIs). Although some studies have reported the genome characteristics of *S. epidermidis*, a comprehensive understanding of its pathogenicity and the genome of ocular trauma-sourced strains is still lacking. In this study, we sequenced, analyzed and reported the whole genomes of 11 ocular trauma-sourced samples of *S. epidermidis* that caused traumatic endophthalmitis. By integrating publicly available genomes, we obtained a total of 187 *S. epidermidis* samples from healthy and diseased eyes, skin, respiratory tract and blood. Combined with pangenome, phylogenetic and comparative genomic analyses, our study supported that *S. epidermidis*, regardless of niche source, exhibits two founder lineages with different pathogenicities. Moreover, we identified potential biomarkers associated with the virulence of *S. epidermidis*, namely, *essD*, *uhpt*, *sdrF*, *sdrG*, *fbe* and *icaABCDR*. The *essD* and *uhpt* genes have high homology with *esaD* and *hpt* in *Staphylococcus aureus*, showing that the genomes of *S. epidermidis* and *S. aureus* may have communicated during evolution, while the *sdrF*, *sdrG*, *fbe*, and *icaABCDR* genes are related to biofilm formation. Compared to *S. epidermidis* from blood sources, ocular-sourced strains causing intraocular infection had no direct relationship with biofilm formation. In conclusion, this study not only provided

additional data resources for studies on *S. epidermidis*, but also improved understanding of the evolution and pathogenicity of different source strains.

Importants In this study, we comprehensively analysed and reported the whole genome sequence (WGS) of the strains that caused traumatic endophthalmitis. Through comparative genomic analyses among 187 *S. epidermidis* samples from healthy and diseased eyes, skin, respiratory tract and blood, we identified 10 potential biomarkers of pathogenic strains. Finally, we revealed *S. epidermidis*-induced traumatic endophthalmitis may be independent of biofilm formation. Overall, our data may facilitate comparative research of *S. epidermidis* and provide clinical value for identifying pathogenic or contaminating strains.

Keywords *Staphylococcus epidermidis*, ocular, whole genome, pan-genomic, evolution, comparative genomic, virulence gene, antimicrobial resistance

Introduction Coagulase-negative *staphylococci* (CoNS) usually live on human skin, can be cultivated from every individual, even healthy persons, and represent one of the major nosocomial pathogens(1). As recorded within the U.S. nationwide Surveillance and Control of Pathogens of Epidemiological Importance (SCOPE) database, the most common isolate recovered from nosocomial bloodstream infections was CoNS (31%) at the end of a 7-year period(2). Additionally, CoNS preferentially colonized the nasal cavity(97%)(3) and ocular surface (60%)(4)(5). *Staphylococcus epidermidis* is the most common CoNS(6), and it plays a central role in the skin microbiota; for example, it can

protect against colonization by skin pathogens(7)(8), maintain the ecological balance of human skin flora(9) and modulate the immune system(10)(11). Once it breaches the skin surface and enters the bloodstream, however, it is considered pathogenic. As the second most common cause of nosocomial infections(12), *S. epidermidis* not only accounts substantially for foreign body-related infections(13) but also causes many eye infections(14)(15)(16)(17), such as bacterial endophthalmitis, keratitis, conjunctivitis and blepharitis, mainly because of ocular trauma(17), ocular surgery, ocular inflammation and the use of contact lenses(14)(17). The data obtained in different ways all showed that CoNS, including *S. epidermidis* (more than 30%), was the most common pathogen causing traumatic endophthalmitis, and more than 50% of patients could not recover effective visual acuity (0.1), among which more than 20% completely lost visual acuity. The different roles of *S. epidermidis* in health and disease give it an important and central role in the human microbiota. Although a high positive rate of *S. epidermidis* was detected in clinical specimens, whether they represent true infection or only colonization/contamination remains to be discussed.

Although much less is known regarding the potential of *S. epidermidis* to cause outbreaks, the advantage of CoNS isolates showed (multiple) resistance to antibiotics (18)(19)(20), as well as their ability to produce biofilms(21)(22), strongly suggesting that modern medicine facilitated the selection process, mainly by the (over)use of antibiotics and the insertion of foreign body devices(13). The biofilm formation and antibiotic

resistance of *S. epidermidis* contributed to the occurrence and persistence of clinical infections(23). Studies conducted worldwide have shown that in clinical specimens, especially in medical device specimens, the isolation rate of *S. epidermidis* is very high(24)(25)(26). *S. epidermidis* forms biofilms on medical devices, such as contact lenses, catheters, and artificial heart valves(27). The detachment of bacterial cells from the biofilm on these devices can lead to bacteremia, increasing morbidity and potential mortality(27). The biofilm formation of *S. epidermidis* was caused by initial attachment by surface proteins (staphylococcal surface proteins *Ssp-1*, *Ssp-2*, Bap homolog protein *Bhp*, autolysin E *AtlE*) and indirect binding, which was accomplished by identifying the microbial surface components of adhesion matrix molecules (MSCRAMMs), such as fibrinogen-binding protein (*SdrG*, *Fbe*) or polysaccharide intracellular adhesion (*ica* *Luci*)(28). It is well known that biofilms are resistant to antibacterial drugs(27). Antibiotic resistance significantly complicates treatment and increases medical costs(29)(30). Mobile genetic elements, such as multidrug resistance binding plasmids and the set of recombinase genes (*ccr*) called staphylococcal chromosome cassette *mec* (*SCCmec*)(31), enable frequent transfer of β -lactam resistance and adaptation to antibiotic selection pressure by *S. epidermidis*(32)(33). The gene *mecA* is present on *SCCmec* and encodes the penicillin-binding protein PBP2a(34). The alteration in PBP2a may confer resistance to methicillin(35). Although methicillin is not used to treat eye infections, it is known that increasing resistance to methicillin can significantly promote

the spread and persistence of multidrug-resistant strains in specific environments

(36)(37)(38).

Recently, with the development of high-throughput sequencing technology, the genome of *S. epidermidis* has also been extensively studied. Conlan et al first analyzed the full pangenome of *S. epidermidis*, including both commensal and nosocomial isolates(27). Sharma et al reported the presence of diverse lineages of *S. epidermidis* isolates in healthy individuals from two geographically diverse locations in India and North America(39). Su et al selected isolates from human and environmental sources from databases to identify genomic determinants associated with the diversity and of *S. epidermidis* strains and their adaptation to various environments(40). Zhou et al characterized the evolutionary trajectory and functional distribution of *S. epidermidis*(41). Microorganisms can be shaped by different host-specific factors, such as disease and health status, however, because *S. epidermidis* is an opportunistic pathogen with a high isolate rate, we remained unsure whether the detected strains were true infection or only contamination, especially in the case of respiratory tract samples, which were generally considered to be contaminated and were not reported. Thus, the pathogenic condition of *S. epidermidis* from different host health statuses should be fully understood. Moreover, although Kirstahler et al reported six whole genomes of *S. epidermidis* from vitreous humor(42), a comprehensive understanding of the genome of ocular strains, particularly ocular trauma-sourced strains, was still lacking, and genetic differences in strain sources

from different host niches remained unclear.

In this study, we sequenced the whole genomes of 11 ocular trauma-sourced *S. epidermidis* isolates. Through incorporating publicly available genomes, we finally obtained 187 *S. epidermidis* isolates from eyes, skin, respiratory tract and blood of healthy and diseased hosts. Through whole-genome sequence analysis and pangenome analysis, we performed a detailed analysis of the host health status diversity and within-individual sources of *S. epidermidis*, and we comprehensively analyzed and reported the whole-genome sequence (WGS) of *S. epidermidis* that causes traumatic endophthalmitis. We used phylogenetic analysis and comparative genomics to reveal the evolutionary relationships of different sources, exploring how different host niches may shape the genetic diversity of *S. epidermidis*. Our study revealed a marked association between evolutionary lineage and host health states. Regardless of niche source, *S. epidermidis* showed two founder lineages that had different pathogenicity. We also identified biomarkers related to *S. epidermidis* pathogenicity. Compared to blood-sourced *S. epidermidis*, traumatic endophthalmitis strains carried different virulence genes, and causing intraocular infection may be independent of biofilm formation. Overall, our study revealed the genetic diversity and pathogenicity of *S. epidermidis* and accomplished a comprehensive comparative analysis of different source genomes.

Materials and methods

Strains. This study analyzed the whole genomes of 187 *Staphylococcus epidermidis* isolates, including 11 ocular strain isolates from the Department of Laboratories, Eye Hospital of Wenzhou Medical University, Wenzhou, China, and 176 available genome sequences downloaded from the GenBank database(43) of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>). The 187 *S. epidermidis* strains were selected to represent known diversity within isolation sources representing human and host health states. For strain sources, we collected 17 ocular *S. epidermidis* strains, 46 blood-sourced strains, 22 respiratory strains, 46 *S. epidermidis* skin-sourced strains, and 46 clinical strains with unknown sources of host niches classified as group clinics. of the remaining *S. epidermidis* strains belonged to the “Others” group. The “Respiratory” group consists of isolates from the nares, lungs, pharyngeal exudate, sputum and bronchoalveolar lavage. Catheter and oral isolates and the reference genomes of strains ATCC14990 and RP62A were classified in the “Others” group. A total of 19 and 75 strains had a definite health or disease state, respectively.

Whole-Genome Sequencing, Assembly, Gene Predictions and Functional

Annotations. The 11 *S. epidermidis* isolates collected from ocular sources were cultured in 5 ml of brain heart infusion broth (BHI) +5% fetal bovine serum (FBS, Gibco) and shaken at 250 rpm and 37°C for 12–16 h. DNA was extracted using a GENErax Bacterial Genome DNA Mini Kit (GENErax, Shanghai, China) following the manufacturer’s protocol. Whole-genome sequencing was performed by Berry Genomics Co., Ltd,

Beijing, China using PacBio SMRT Technology. SMRTbell libraries were prepared using the SMRTbell Express Template Prep Kit 2.0 (PacBio kit). Long read data were assembled by canu 2.1.1(44), and assembly polisher was used with pbmm2 (v1.4.0) and gcpp (v2.0.0) via the bioconda package. Quast (v5.0.2)(45), checkM (v1.1.3)(46) and busco (v5.0.0)(47) were used to assess the quality of assembled genomes. The genes of all genomes, including the 11 ocular *S. epidermidis* genomes that we sequenced and obtained from NCBI, were predicted and annotated by Prokka (v1.13)(48).

Pan-Genome Analysis. Pangenome analysis was carried out for all 187 genomes and isolate genomes from different niches by BPGA v1.3(49) using default parameters and 50% identity as the cutoff. The annotation files generated by Prokka were provided to BPGA as an input. The COGs of the core sequence, accessory sequence and unique sequence after Pangenome analysis were further verified by performing functional annotations in the EggNOG Database (v5.0)(50) via eggno-mapper (v2.0)(51) with the Diamond parameter.

Characterization of Strains. kSNP3.1(52) was used to build the phylogenetic tree, a validated method without alignment by 19 k-mer length, and a total of 8760 core SNPs were identified. Then, a maximum-likelihood core SNP tree was constructed by RAxML (v8.2.12)(53) with 100 bootstraps. The phylogenetic tree was visualized and beautified by online itol (v6.0) (<https://itol.embl.de/>)(54).

Multilocus sequence typing (MLST) of 187 *S. epidermidis* strains was performed with the MLST 2.0 online server (<https://cge.cbs.dtu.dk/services/MLST/>)(55). Online SCCmecFinder 1.2 (<https://cge.cbs.dtu.dk/services/SCCmecFinder/>) was used to identify SCCmec elements in sequenced *S. epidermidis* isolates.

ResFinder(56) and Resistance Gene Identifier (RGI) of the Comprehensive Antibiotic Resistance Database (CARD)(57) were employed to identify antimicrobial resistance genes (AMRs). The resistance genes were verified by the paper diffusion method (MHA) and broth dilution method. The BLASTp program was used to search all protein sequences of 187 *S. epidermidis* strains against the Virulence Factor Database (VFDB) (<http://www.mgc.ac.cn/VFs/>)(58). Compared with the virulence genes in the database at an e-value < 1e-10, only query genes with an identity higher than 40% and a coverage higher than 70% were considered potential virulence genes(59). Virulence factor functional annotations were based on the categories and subcategories presented in VFDB. An alignment for each of the extracted candidate virulence determinant genes was constructed using Clustal Omega(60) and visualized by ESPript(61).

Statistical Analyses. The significance of core gene and accessory gene abundance in COG categories was examined using Fisher's exact test. The disease-associated and ocular-associated predicted accessory genes with known function and annotated virulence genes were analyzed using Fisher's exact tests and the FDR correction of P values. All

statistical analyses were carried out using the R package (version: 4.0.2). A P-value of < 0.05 was regarded as statistically significant.

Data availability. The sample and sequence data obtained in this study have been submitted to the NCBI BioSample and Sequence Read Archive (SRA) under BioProject accession number PRJNA753005.

Results

General Features of Sequenced Ocular Isolate Genomes. Eleven *Staphylococcus epidermidis* isolates from *S. epidermidis*-infected ocular endophthalmitis were sequenced and assembled. We used pbmm2 to map the draft genomes to raw subreads and ultimately counted the coverage depth. High-quality data were obtained with a mean coverage depth >5000X and number of contigs in the range of 1 – 6. Annotation of these isolates revealed predicted CDSs (from coding sequences) ranging from 2,247 to 2,692.

The sequencing and annotation details of the 11 ocular isolates are listed in **Table 1**. The 11 draft genomes were 2.58 ± 0.1 Mb in size and 32.14% GC on average. It seemed that the features of sequenced ocular isolate genomes were not significantly different from those obtained from skin and blood. We chose a representative assembled ocular strain genome and drew a genomics map with Circos(62) to illustrate the genome characterization (**Figure 1**).

Pan-Genome and Functional Characterization of *S. epidermidis* from Different

Sources. Although many studies have shown that pangenomes characterize *S.*

epidermidis(27), core and pangenome features of ocular strains are lacking. Thus, to clarify the pangenomic characteristics of *S. epidermidis* from ocular sources and discrepancies among different sources, we determined general genetic similarities and differences within ocular *S. epidermidis* and all 187 strains. Gene accumulation curves(63) (**Figure 2A and Figure 2B**) showed that the number of core genomes fits an exponential decay curve that plateaus at 749 and 1,931 genes, respectively, while the pangenome data fit a power law curve ($y=a \cdot x^b$), indicating an open pangenome where each genome sequence added a number of new genes, as reported(27). To a certain degree, the capacity of acquiring exogenous DNA of the organism partially determines the pangenome state (“open” or “close”)(64), especially for species living in bacterial communities, such as skin inhabitants.

The function of the genes within the pangenome of whole collected and different source strains was investigated by assigning all gene clusters to clusters of orthologous groups (COGs)(65) (**Figure 2C**). For entirety, the results revealed that there were a total of 685/749 core genes, 1,792/2,783 accessory genes and 234/468 unique genes among all 187 *S. epidermidis* strains annotated to COG categories. Specifically, by Fisher’s exact test ($FDR < 0.05$), 5 of 22 COG categories were significantly enriched in core genes and almost associated with metabolism and biogenesis: energy production and conversion, nucleotide transport and metabolism, coenzyme transport and metabolism, translation, ribosomal structure and biogenesis and inorganic ion transport and metabolism.

Additionally, while replication, recombination and repair were enriched in both accessory and unique genes, secondary metabolite biosynthesis, transport and catabolism and defense mechanisms were enriched in unique genes. Second, we compared the enrichment of COG function among strains from different sources, including the blood, eyes, respiratory tract, and skin, and showed similar results: core genes related to metabolism and biogenesis. However, the difference from the above is that defense mechanisms were enriched in accessory genes of all strains except eye-sourced ones. In addition, transcription was significantly enriched in accessory genes of blood-sourced strains, while other sources were enriched in unique genes. In summary, compared to the core genes enriched in biogenesis and metabolism, the enrichment of replication, recombination and repair, defense mechanisms, and transcription among accessory genes and unique genes was driven by diversity in recombinase and integrase, ABC-type multidrug transporters, and transcriptional regulators, respectively. These abundant genes often appear to transfer horizontally between strains, leading to the spread of virulence and resistance genes between strains, which affects the pathogenicity of bacteria(66).

Phylogenetic Relationship and Associated Typing among *S. epidermidis*. To infer the phylogenetic relationship of the *S. epidermidis* strains from different sources, we used 8760 core SNPs to build a single-nucleotide polymorphism-based phylogenetic tree. As expected, the 187 *S. epidermidis* isolates formed two distinct groups termed I and II, as previously reported(27)(41) (**Figure 2D**), suggesting the presence of two founder

lineages. Moreover, we found that ST2 and ST5 strains were present in Cluster II and correlated with definite disease hosts or clinical strains, while ST691 strains were present in Cluster I and were related to healthy skin with the same genetic distance.

It is worth mentioning that although the ocular isolates we examined are varied in phylogeny, they all had a close distance to different sources from diseased hosts in both clades, which proved that the ocular strains we collected were pathogenic. However, the ST typing of ocular strains was highly diverse and included rare specimens: 4/11 strains we collected were unknown.

S. epidermidis contains staphylococcal chromosome cassette *mec* (SCC*mec*), called methicillin-resistant *S. epidermidis* (MRSE)(27). SCC*mec* is a mobile genetic element defined by combinations of *mec* gene complexes, cassette recombinases and accessory genes and carries the central determinant for broad-spectrum beta-lactam resistance encoded by the *mecA* gene, a mobile genetic element of *Staphylococcus* species(27)(67). We predicted the SCC*mec* element (SCC*mec* type) and found that it existed in 94/187 strains. Interestingly, almost all SCC*mec*-positive strains (96.8%, 91/94) were in Cluster II, indicating that compared to SCC*mec*-negative strains, *S. epidermidis* carrying the SCC*mec* element may be more pathogenic.

The Virulence Characteristics of *S. epidermidis*. Although *S. epidermidis* harbors few classical virulence determinants(6), it is an opportunistic pathogen with many virulence factors. To analyze the different pathogenic potentials within *S. epidermidis* from

different sources, 99 genetic loci were identified to be related to virulence based on the virulence factor database (VFDB), which was grouped into 18 categories (**Figure 4**), including 21 (21/99, 21.2%) virulence genes that coexisted in all *S. epidermidis* genomes. Surprisingly, we identified several virulence genes, *icaABCD*, *hpt* and *esaD*, that were significantly associated with *S. epidermidis* from diseased sources (Fisher's exact test, FDR < 0.01). Polysaccharide intercellular adhesion (*icaABCD*) genes that encode biofilm-associated genes for poly-N-acetylglucosamine synthesis were found in 48% (36/75) of the diseased isolates; notably, only 1 of 19 healthy strains contained the *icaABCD* gene, which disagreed with previous studies that found this gene in 60% of commensal isolates. By pairwise comparison of the enrichment of virulence gene in ocular-sourced strains and strains from other niches, we found 30 *icaABCD* genes in 46 blood-sourced strains and 3 in 17 ocular-sourced strains. We also found the Ser-Asp-rich proteins *sdrE* to be enriched in ocular-sourced strains (16/17) compared to blood-sourced strains (24/46) (Fisher's exact test, FDR < 0.05). Two toxin genes, *sell* and *sec*, which were present in only two strains of blood-sourced *S. epidermidis* (SE90 and SE95), as reported previously(68), were also identified in two ocular strains, combining the results of phylogenetics showing that the strains had a close evolutionary distance, implying that they may have the similar founder lineages.

Gene Differences between Healthy and Disease Source Genomes. *Staphylococcus*

epidermidis is a coagulase-negative and gram-positive *staphylococcus* that is part of the

skin, mucosa microflora(69). They also had high relative abundances and high positivity rates on the ocular surface. It is the second leading cause of nosocomial infections(70). Within the 187 strain genomes we analyzed, 75 were definitely from diseased hosts and 19 were from healthy individuals. To investigate whether the presence of specific genes was significantly correlated with different sources of strains, including isolates from different healthy hosts. We devoted to accessory genes, which were defined to be gene clusters neither present in all genomes nor exit in only a single genome. Distinctly, the genes were divided into two clusters, which strictly differentiated the disease group and healthy isolates (Fisher's exact test, FDR < 0.05) (**Figure 4A**). Ulteriorly, the two groups of genes were functionally classified by KEGG function (**Figure 4B**). Both congruently, metabolism accounted for the largest proportion, followed by signaling and cellular processes. Comparing the two groups in the subcategory of KEGG function, disease-associated genes had a larger proportion involved in amino acid metabolism, antimicrobial resistance and transcription factors, while carbohydrate metabolism had a high ratio of health-related genes.

Among the differentially expressed genes between strains from diseased and healthy sources, we found some key disease-related genes, including the virulence genes *essD*, *uhpt*, *sdrF*, *sdrG*, *fbe*, and *icaABCDR* and the transcriptional regulators *lrpC* and *cysL* (**Figure 5A**). *SdrF*, *sdrG* and *fbe*, which are microbial surface components recognizing adhesive matrix molecules (MSCRAMMs), showed homology with the virulence gene

sdrE, but not all of the gene sequences were identified as virulence factors. By multiple alignment with Clustal omega, we found that compared to both the virulence gene *sdrG* protein sequences and *Staphylococcus aureus*, nonvirulence gene *sdrG* protein sequences mutated in the β -strand (**Figure 5C**), which may decrease virulence.

Interestingly, to compare genes between ocular and other places of organism, we found that biofilm formation-related genes *icaABCD* and *pgaC* were markedly enriched in blood but not in ocular tissue (**Figure 5B**). Moreover, although *sdrG* seemingly enriched in ocular cells, the β -strand mutation rate reached 42.86% (6/14), which may lead to weaker adherence ability of the microbes to components of the extracellular matrix of the host, suggesting that biofilm formation may not be a direct factor for intraocular infection of *S. epidermidis*.

Antimicrobial Resistance across *S. epidermidis*. Antimicrobial resistance is very common among *S. epidermidis* isolates, and genes within *S. epidermidis* involved in antibiotic resistance contribute to the persistence of clinical infection and often limit treatment options(71). To investigate antimicrobial resistance across *S. epidermidis*, we analyzed all known AMR genes within our 187 genomic datasets. According to our analysis of the ResFinder and CARD databases, we found 41 different genes involved in resistance to 19 antibiotics (**Figure 6**). Nearly all isolates carried at least three antibiotic resistance genes. Among the genes involved in AMR, our data showed that two genes, *norA*, which is associated with fluoroquinolone antibiotics, from the AMR gene family

with major facilitator superfamily (MFS) antibiotic efflux pumps and *dfrC*, a diaminopyrimidine antibiotic-associated gene, were conserved in all strains. This result was consistent with the data in the comprehensive antibiotic resistance database (CADR: <https://card.mcmaster.ca/ontology>).

Based on Fisher's exact test of strains from different host health states and niches, we found that the methicillin resistance genes *mecA* and *mecRI* (FDR <0.01), fluoroquinolone antibiotic gene *ace* (FDR <0.05) and aminoglycoside antibiotic gene *aac(6')-Ie-aph(2'')-Ia* (FDR <0.001) were enriched in strains isolated from diseased hosts. The presence of *mecA* was consistent with the SCC*mec*-positive strains because the mobile genetic element SCC*mec* carried the gene *mecA*. Antibiotic susceptibility testing currently uses oxacillin and cefoxitin to assess resistance to methicillin, and the results showed that the intraocular strain oxacillin resistance rate was 5/11 (Table 2), which was consistent with the results for cefoxitin and *mecA* gene carriers. Moreover, strains from healthy skin had significant enrichment of the *msrA* and *mgrA* genes with multidrug resistance (FDR <0.001), while isolates from the ocular, blood and respiratory tracts had no significantly enriched antibiotic resistance genes, which suggested that regardless of niche sources, none of these isolates were resistant to specific antibiotics at the genetic level.

Discussion Coagulase-negative *Staphylococcus* was the most common pathogen in traumatic endophthalmitis(72). *S. epidermidis*, which colonizes the normal mucosa, skin

flora and intraocular tissue of humans and other mammals, is the most common coagulase-negative gram-positive *Staphylococcus*(69)and one of the major leading causes of clinical infections. What still needs to be discussed is whether *S. epidermidis* is isolated as a pathogen or contaminant. Although the genome characteristics of *S. epidermidis* have been studied in recent years, a comprehensive understanding of the genomes of ocular-sourced strains is lacking. To the best of our knowledge, this is the first and largest collection of bacterial genome sequences isolated from patients with *S. epidermidis* intraocularly. By collecting, sequencing and analyzing the genomes of eleven intraocular isolates and incorporating publicly available genomes for *S. epidermidis*, we gained insight into the phylogenetic and molecular characteristics of intraocular and other niche pathogens. The genetic differences between pathogenic and commensal *S. epidermidis* were also investigated comprehensively.

In this study, we compared the phylogenetic diversity and genome characteristics of *S. epidermidis* from different niches and different host health states. The host niches, including the eye, blood, skin and respiratory tract, , were all from diseased hosts. The 11 ocular-sourced genomes we sequenced revealed similar high-quality benchmark data, including genome size, GC content, and the number of predicted genes, which were similar to the data we collected from public databases and had a relatively compact genome with an average size of approximately 2.58 Mb. Furthermore, the completeness of each whole-genome sequencing exceeded 99%, and the contamination level of the

sequencing library was also very low. The high coverage of genome assembly helped us obtain a complete and accurate genome. Consistent with the results of previous studies (27)(39)(41), whether in the pangenome analysis of all 187 strains or strains of various niches, including ocular sources, all showed that the size of the *S. epidermidis* genome was relatively constant, extracted from a larger gene pool, indicating an increase in the “open” pangenome; with each genome sequence, some new genes were added. To some extent, the pangenomic state ("open" or "closed") of an organism depends in part on its ability to obtain exogenous DNA(64); for example, the large number of genes involved in the mobile group makes horizontal gene transfer between staphylococcal strains easier, and mobile genetic elements such as *SCCmec*, *ACME*, and plasmids lead to an increase in the "open" pangenome(73). The phylogenetic tree constructed based on genome-wide core SNPs reveals important details that were not available in the use of traditional single gene markers (16S rDNA) or multilocus sequence typing (MLST)(40). From the analysis of the phylogenetic tree, we found that 187 strains of *S. epidermidis* formed two distinct clusters with different pathogenic abilities, and clinically pathogenic strains were generally ST5 and ST2, while the ST691 strains were derived from healthy skin. The earlier reported phenomenon of evolutionary distance between ST2 *S. epidermidis* was extremely short(40) and was also observed in the ST5 and ST691 strains.

It is well established that *S. epidermidis* is a common human skin commensal; if it breaks through the surface of the skin and enters the blood, it is considered a

pathogen(12). Similarly, *S. epidermidis* is an important commensal on the ocular surface(74), but usually caused by trauma, it can enter the eyes to cause intraocular infection. Through whole-genome analysis, we have provided a powerful framework to redefine species clustering in the genus, locate genetic traits, and rate the importance of disease-causing genes based on their presence or absence. Using comparative genomics methods to analyze the genomes of *S. epidermidis* from different sources, we obtained the potential pathogenic marker genes *lrpC*, *cysL*, *essD*, *uhpt*, *sdrF*, *sdrG*, *fbe*, and *icaABCDR*. *LrpC* and *cysL* both have helix-turn-helix (HTH)-type transcriptional regulators. The *lrp*-like regulatory factor consisted of a helix-turn-helix (HTH)-type n-terminal DNA binding domain, which was connected to the C-terminal RAM domain (amino acid metabolism regulation) and was responsible for cofactor binding and oligomerization(75). Thus, *lrpC* was a transcriptional regulator with a possible role in the regulation of amino acid metabolism and the growth phase transition. *CysL* belongs to the *lysR* family transcriptional regulator. The *lysR* HTH domain is a DNA-binding, winged helix-turn-helix (wHTH) domain consisting of approximately 60 residues in the *lysR*-type transcription regulator (LTTR). LTTR is one of the most common regulatory factor families in prokaryotes. The c-terminus of the *lysR* protein can contain a regulatory domain with two subdomains, which participate in (1) coinducer recognition/reaction and (2) DNA binding and response(76). LTTRs can activate the transcription of operons and regulons involved in the regulation of various functions, such as amino acid biosynthesis,

CO₂ fixation, antibiotic resistance, virulence factor regulation, nitrogen-fixing bacterial nodulation, oxidative stress response or aromatic compound catabolism. However, the specific functions of the transcriptional regulators *cysL* and *lrpC* in *S. epidermidis* need to be further studied.

In this work, pathogenic marker genes *essD*, *uhpt*, *sdrF*, *sdrG*, *fbe*, and *icaABCD* were identified as virulence genes. Interestingly, we found that the potential virulence genes *essD* and *uhpt* had high homology with *esaD* and *hpt* in *Staphylococcus aureus*. *EsaD* is a type VII secretion system secreted protein, a nuclease toxin, which may play a key role in bacterial competition(77). Hexose phosphate is an important carbon source in the cytoplasm of host cells(78). Bacterial pathogens invade, survive and breed in different host epithelial cells, utilizing hexose phosphate in the host cytoplasm to obtain energy and synthesize cell components through the hexose phosphate transport (HPT) system (78). The HPT system of *S. aureus*, which includes the *hptRS* (a new type of two-component regulatory system), *hptA* (a phosphate sensor) and *uhpT* (a hexose phosphate transporter) genes, made it stable in the host cell and may be an important target for the development of new anti-staphylococcal therapies(78). The potential virulence genes *essD* and *uhpt* of *S. epidermidis* identified in this study could be due to horizontal gene transfer between *Staphylococcus* strains; for example, the genomes of *S. epidermidis* and *S. aureus* exchanged during evolution. Furthermore, the genes *sdrF*, *sdrG*, *fbe*, and polysaccharide intercellular adhesion gene (*icaABCD*) were considered to

be related to biofilm formation(28). The formation of biofilms was the main virulence factor *S. epidermidis* contributed to the persistence of clinical infections(23). Here, the seven genes all pertain to adhesive molecules, which are well-known factors involved in biofilm formation. *SdrF*, *sdrG* and *fbe* are a subset of microbial surface components recognizing adhesive matrix molecules (MSCRAMMs)(28); they were covalently anchored to the cell wall and characterized by a segment composed of repeated serine aspartate (SD) dipeptides(79). MSCRAMMs are bacterial surface proteins that mediate the adhesion of microorganisms to the host's extracellular matrix components(79). By sequence analysis, we found that not all of the *sdr* gene sequences were identified as virulence factors. Through further multiple sequence alignment, we found that the virulence sequence of *sdrF* was very different from the nonvirulence sequence, including the length of the sequence, while only the β chain mutation occurred between the two *sdrG* genes. We suspected that the occurrence of these mutations may reduce the virulence of *sdrG*, and it was clear that mutagenesis is overrepresented in intraocular isolates compared to blood isolates. According to our enrichment analysis, we found it was possible to differentiate the intralocular pathogenic strains from those of blood, with biomarkers related to biofilm formation polysaccharide intercellular adhesion *icaABCD* enriched in blood but not in ocular tissue. This may be related to the fact that *S. epidermidis* was more likely to form biofilms on medical devices such as catheters and artificial heart valves(2). Bacterial cells on these devices can break away from the biofilm

and enter the bloodstream, leading to bacteremia, increasing morbidity and potential mortality(2). On the other hand, this also reflected that the intraocular infection of *S. epidermidis* was not directly related to its biofilm formation. Of course, a larger amount of data and further experiments are required to prove this point.

S. epidermidis has also been found to be a treasure trove of antibiotic resistance(27). Through a rare horizontal gene transfer event, its determinants of toxicity were shared with other more pathogenic species, such as *S. aureus*, as confirmed in previous studies(80). In our study, we found that 94/187 (50.2%) of the collected *S. epidermidis* species were methicillin-resistant *Staphylococcus epidermidis* (MRSE) with the *mecA* gene and SCC*mec* elements. In particular, the SCC*mec* cassette that conferred β -lactam resistance was often transferred between staphylococcal strains, enabling them to rapidly evolve and adapt to antibiotic selection pressures and providing additional competitive advantages. This provided strong support for the idea that pathogens influence the risk of infection by the background microbiota through HGT of the pathogenic host, thereby increasing the risk of infection in other parts of the body, such as methicillin-resistant *S. aureus* in the nasal cavity affecting *S. epidermidis*-infected endophthalmitis(40).

Conclusion Our study provided information on the molecular characteristics of different pathogenic *Staphylococcus epidermidis* isolated from different host niches around the world, including ocular strains that were overlooked in previous studies. Pangenome and phylogenetic analysis that *S. epidermidis* had an open pangenome and two founder

lineages that had different pathogenicities. Interestingly, methicillin-resistant *S. epidermidis* was concentrated in the pathogenic branch. Although the endophthalmitis-associated *S. epidermidis* isolated in this study was relatively dispersed in evolution, they were close to the clinical pathogenic strains, which proved their pathogenicity. Based on comparative genomics, we identified 8 potential biomarkers related to strain pathogenicity and provided evidence that horizontal gene transfer (HGT) may occur between *Staphylococcus* strains. Moreover, we reported the complete genome sequence of *S. epidermidis* that caused traumatic endophthalmitis and found that those strains causing intraocular infection may be independent of biofilm formation. Overall, this study revealed genetic diversity and pathogenic differences in different sources of *S. epidermidis*.

References

1. Piette A., Verschraegen G.(2009). Role of coagulase-negative staphylococci in human disease. *Vet Microbiol*, 134(null), 45-54. doi:10.1016/j.vetmic.2008.09.009.
2. Wisplinghoff Hilmar., Bischoff Tammy., Tallent Sandra M., Seifert Harald., Wenzel Richard P., Edmond Michael B.(2004). Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis*, 39(3), 309-17. doi:10.1086/421946.
3. Wos-Oxley Melissa L., Plumeier Iris., von Eiff Christof., Taudien Stefan., Platzer Matthias., Vilchez-Vargas Ramiro., Becker Karsten., Pieper Dietmar H.(2010). A poke

into the diversity and associations within human anterior nares microbial communities.

ISME J, 4(7), 839-51. doi:10.1038/ismej.2010.15.

4. Graham Joanna E., Moore Jonathan E., Jiru Xu., Moore John E., Goodall Edward A.,

Dooley James S G., Hayes Velma E A., Dartt Darlene A., Downes C Stephen., Moore

Tara C B.(2007). Ocular pathogen or commensal: a PCR-based study of surface bacterial

flora in normal and dry eyes. Invest Ophthalmol Vis Sci, 48(12), 5616-23.

doi:10.1167/iovs.07-0588.

5. Willcox MD. (2013). Characterization of the normal microbiota of the ocular surface.

Exp. Eye Res. 117:99 – 105. doi:10.1016/j.exer.2013.06.003.

6. Otto Michael.(2004). Virulence factors of the coagulase-negative staphylococci. Front

Biosci, 9(undefined), 841-63. doi:10.2741/1295.

7. Cogen Anna L., Yamasaki Kenshi., Sanchez Katheryn M., Dorschner Robert A., Lai

Yuping., MacLeod Daniel T., Torpey Justin W., Otto Michael., Nizet Victor., Kim Judy E.,

Gallo Richard L.(2010). Selective antimicrobial action is provided by phenol-soluble

modulins derived from Staphylococcus epidermidis, a normal resident of the skin. J

Invest Dermatol, 130(1), 192-200. doi:10.1038/jid.2009.243.

8. Cogen Anna L., Yamasaki Kenshi., Muto Jun., Sanchez Katheryn M., Crotty

Alexander Laura., Tanios Jackelyn., Lai Yuping., Kim Judy E., Nizet Victor., Gallo

Richard L.(2010). Staphylococcus epidermidis antimicrobial delta-toxin (phenol-soluble

- modulin-gamma) cooperates with host antimicrobial peptides to kill group A
Streptococcus. PLoS One, 5(1), e8557. doi:10.1371/journal.pone.0008557.
9. Schommer Nina N., Gallo Richard L.(2013). Structure and function of the human skin
microbiome. Trends Microbiol, 21(12), 660-8. doi:10.1016/j.tim.2013.10.001.
10. Egert M., Simmering R., Riedel C U.(2017). The Association of the Skin Microbiota
With Health, Immunity, and Disease. Clin Pharmacol Ther, 102(1), 62-69.
doi:10.1002/cpt.698.
11. Linehan Jonathan L., Harrison Oliver J., Han Seong-Ji., Byrd Allyson L.,
Vujkovic-Cvijin Ivan., Villarino Alejandro V., Sen Shurjo K., Shaik Jahangheer.,
Smelkinson Margery., Tamoutounour Samira., Collins Nicholas., Bouladoux Nicolas.,
Dzutsev Amiran., Rosshart Stephan P., Arbuckle Jesse H., Wang Chyung-Ru., Kristie
Thomas M., Rehmann Barbara., Trinchieri Giorgio., Brenchley Jason M., O'Shea John
J., Belkaid Yasmine.(2018). Non-classical Immunity Controls Microbiota Impact on Skin
Immunity and Tissue Repair. Cell, 172(4), 784-796.e18. doi:10.1016/j.cell.2017.12.033.
12. Otto Michael.(2009). Staphylococcus epidermidis--the 'accidental' pathogen. Nat Rev
Microbiol, 7(8), 555-67. doi:10.1038/nrmicro2182.
13. Karsten Becker, Christine Heilmann, Georg Peters. (2014) Coagulase-Negative
Staphylococci. Clinical Microbiology Reviews, 870 – 926.doi:10.1128/CMR.00109-13.

14. Park Young Min., Kwon Han Jo., Lee Jong Soo.(2015). Microbiological study of therapeutic soft contact lenses used in the treatment of recurrent corneal erosion syndrome. *Eye Contact Lens*, 41(2), 84-6. doi:10.1097/ICL.0000000000000068.
15. Bispo Paulo J M., Hofling-Lima Ana Luisa., Pignatari Antonio C C.(2014). Characterization of ocular methicillin-resistant *Staphylococcus epidermidis* isolates belonging predominantly to clonal complex 2 subcluster II. *J Clin Microbiol*, 52(5), 1412-7. doi:10.1128/JCM.03098-13.
16. Schimel, A. M., Miller, D., and Flynn, H. W. (2013). Endophthalmitis isolates and antibiotic susceptibilities: a 10-year review of culture-proven cases. *Am. J. Ophthalmol.* 156, e50 – e52. doi: 10.1016/j.ajo.2013.01.027.
17. Keay, L., Edwards, K., Naduvilath, T., Taylor, H. R., Snibson, G. R., Forde, K., et al. (2006). Microbial keratitis: predisposing factors and morbidity. *Ophthalmology* 113, 109 – 116. doi: 10.1016/j.opthta.2005.08.013.
18. Kresken M., Becker K., Seifert H., Leitner E., Körber-Irrgang B., von Eiff C., Löschmann P-A., Study Group.(2011). Resistance trends and in vitro activity of tigecycline and 17 other antimicrobial agents against Gram-positive and Gram-negative organisms, including multidrug-resistant pathogens, in Germany. *Eur J Clin Microbiol Infect Dis*, 30(9), 1095-103. doi:10.1007/s10096-011-1197-y.
19. Ma Xiao Xue., Wang En Hua., Liu Yong., Luo En Jie.(2011). Antibiotic susceptibility of coagulase-negative staphylococci (CoNS): emergence of

- teicoplanin-non-susceptible CoNS strains with inducible resistance to vancomycin. *J Med Microbiol*, 60(null), 1661-1668. doi:10.1099/jmm.0.034066-0.
20. Lepointeur M., Royer G., Bourrel A S., Romain O., Duport C., Doucet-Populaire F., Decousser J-W.(2013). Prevalence of resistance to antiseptics and mupirocin among invasive coagulase-negative staphylococci from very preterm neonates in NICU: the creeping threat? *J Hosp Infect*, 83(4), 333-6. doi:10.1016/j.jhin.2012.11.025.
21. Dimitriou G., Fouzas S., Giormezis N., Giannakopoulos I., Tzifas S., Foka A., Anastassiou D E., Spiliopoulou I., Mantagos S.(2011). Clinical and microbiological profile of persistent coagulase-negative staphylococcal bacteraemia in neonates. *Clin Microbiol Infect*, 17(11), 1684-90. doi:10.1111/j.1469-0691.2011.03489.x.
22. de Silva G D I., Kantzanou M., Justice A., Massey R C., Wilkinson A R., Day N P J., Peacock S J.(2002). The ica operon and biofilm production in coagulase-negative Staphylococci associated with carriage and disease in a neonatal intensive care unit. *J Clin Microbiol*, 40(2), 382-8. doi:10.1128/JCM.40.2.382-388.2002.
23. Schoenfelder Sonja M K., Lange Claudia., Eckart Martin., Hennig Susanne., Kozytska Svitlana., Ziebuhr Wilma.(2010). Success through diversity - how Staphylococcus epidermidis establishes as a nosocomial pathogen. *Int J Med Microbiol*, 300(6), 380-6. doi:10.1016/j.ijmm.2010.04.011.
24. Cherifi S., Byl B., Deplano A., Nonhoff C., Denis O., Hallin M.(2013). Comparative epidemiology of Staphylococcus epidermidis isolates from patients with catheter-related

bacteremia and from healthy volunteers. J Clin Microbiol, 51(5), 1541-7.

doi:10.1128/JCM.03378-12.

25. Ehlers Marthie M., Strasheim Wilhelmina., Lowe Michelle., Ueckermann Veronica.,

Kock Marleen M.(2018). Staphylococcus epidermidisMolecular Epidemiology of

Implicated in Catheter-Related Bloodstream Infections at an Academic Hospital in

Pretoria, South Africa. Front Microbiol, 9(undefined), 417.

doi:10.3389/fmicb.2018.00417.

26. Ardon C B., Prens E P., Fuursted K., Ejaz R N., Shailes J., Jenssen H., Jemec G B

E.(2019). Biofilm production and antibiotic susceptibility of Staphylococcus epidermidis

strains from Hidradenitis Suppurativa lesions. J Eur Acad Dermatol Venereol, 33(1),

170-177. doi:10.1111/jdv.15183.

27. Conlan Sean., Mijares Lilia A., NISC Comparative Sequencing Program., Becker

Jesse., Blakesley Robert W., Bouffard Gerard G., Brooks Shelise., Coleman Holly., Gupta

Jyoti., Gurson Natalie., Park Morgan., Schmidt Brian., Thomas Pamela J., Otto Michael.,

Kong Heidi H., Murray Patrick R., Segre Julia A.(2012). Staphylococcus epidermidis

pan-genome sequence analysis reveals diversity of skin commensal and hospital

infection-associated isolates. Genome Biol, 13(7), R64. doi:10.1186/gb-2012-13-7-r64.

28. Foster Timothy J., Geoghegan Joan A., Ganesh Vannakambadi K., Höök

Magnus.(2014). Adhesion, invasion and evasion: the many functions of the surface

- proteins of *Staphylococcus aureus*. *Nat Rev Microbiol*, 12(1), 49-62.
doi:10.1038/nrmicro3161.
29. Foster Timothy J.(2017). Antibiotic resistance in *Staphylococcus aureus*. *Current status and future prospects*. *FEMS Microbiol Rev*, 41(3), 430-449.
doi:10.1093/femsre/fux007
30. Lee Jean Y H., Monk Ian R., Gonçalves da Silva Anders., Seemann Torsten., Chua Kyra Y L., Kearns Angela., Hill Robert., Woodford Neil., Bartels Mette D., Strommenger Birgit., Laurent Frederic., Dodémont Magali., Deplano Ariane., Patel Robin., Larsen Anders R., Korman Tony M., Stinear Timothy P., Howden Benjamin P.(2018). Global spread of three multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat Microbiol*, 3(10), 1175-1185. doi:10.1038/s41564-018-0230-7.
31. McManus Brenda A., Coleman David C., Deasy Emily C., Brennan Gráinne I., O'Connell Brian., Monecke Stefan., Ehricht Ralf., Leggett Bernadette., Leonard Nola., Shore Anna C.(2015). Comparative Genotypes, *Staphylococcal Cassette Chromosome mec* (SCCmec) Genes and Antimicrobial Resistance amongst *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* Isolates from Infections in Humans and Companion Animals. *PLoS One*, 10(9), e0138079. doi:10.1371/journal.pone.0138079.
32. Miragaia M., Thomas J C., Couto I., Enright M C., de Lencastre H.(2007). Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. *J Bacteriol*, 189(6), 2540-52. doi:10.1128/JB.01484-06.

33. Bloemendaal Alexander L A., Brouwer Ellen C., Fluit Ad C.(2010). Methicillin resistance transfer from *Staphylococcus epidermidis* to methicillin-susceptible *Staphylococcus aureus* in a patient during antibiotic therapy. PLoS One, 5(7), e11841. doi:10.1371/journal.pone.0011841.
34. Katayama Y., Ito T., Hiramatsu K.(2000). A new class of genetic element, *staphylococcus cassette chromosome mec*, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother, 44(6), 1549-55. doi:10.1128/AAC.44.6.1549-1555.2000.
35. Hartman B J., Tomasz A.(1984). Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. J Bacteriol, 158(2), 513-6. doi:10.1128/jb.158.2.513-516.1984.
36. Alekshun, M. N., and Levy, S. B. (2007). Molecular mechanisms of antibacterial multidrug resistance. Cell 128, 1037 – 1050. doi: 10.1016/j.cell.2007.03.004.
37. Asbell, P. A., Sahm, D. F., Shaw, M., Draghi, D. C., and Brown, N. P. (2008). Increasing prevalence of methicillin resistance in serious ocular infections caused by *Staphylococcus aureus* in the United States: 2000 to 2005. J. Cataract. Refract. Surg. 34, 814 – 818. doi: 10.1016/j.jcrs.2008.01.016.
38. Lichtinger Alejandro., Yeung Sonia N., Kim Peter., Amiran Maoz D., Iovieno Alfonso., Elbaz Uri., Ku Judy Y F., Wolff Rachel., Rootman David S., Slomovic Allan

- R.(2012). Shifting trends in bacterial keratitis in Toronto: an 11-year review. Ophthalmology, 119(9), 1785-90. doi:10.1016/j.ophtha.2012.03.031.
39. Sharma Shikha., Chaudhry Vasvi., Kumar Sanjeet., Patil Prabhu B.(2018). Staphylococcus epidermidisPhylogenomic Based Comparative Studies on Indian and American Commensal Isolates. Front Microbiol, 9(undefined), 333. doi:10.3389/fmicb.2018.00333
40. Su Fei., Tian Rui., Yang Yi., Li Hexin., Sun Gaoyuan., Li Ying., Han Bingqing., Xu Xiaomao., Chen Xue., Zhao Gang., Cui Hongyuan., Xu Hongtao.(2020). Staphylococcus epidermidisComparative Genome Analysis Reveals the Molecular Basis of Niche Adaptation of Strains. Front Genet, 11(undefined), 566080. doi:10.3389/fgene.2020.566080
41. Zhou Wei., Spoto Michelle., Hardy Rachel., Guan Changhui., Fleming Elizabeth., Larson Peter J., Brown Joseph S., Oh Julia.(2020). Host-Specific Evolutionary and Transmission Dynamics Shape the Functional Diversification of Staphylococcus epidermidis in Human Skin. Cell, 180(3), 454-470.e18. doi:10.1016/j.cell.2020.01.006.
42. Kirstahler, P., Bjerrum, S. S., Friis-Møller, A., la Cour, M., Aarestrup, F. M., Westh, H., & Pamp, S. J. (2018). Genomics-based identification of microorganisms in human ocular body fluid. Scientific reports, 8(1), 1-14. doi:10.1038/s41598-018-22416-4.

43. Sayers Eric W., Cavanaugh Mark., Clark Karen., Ostell James., Pruitt Kim D., Karsch-Mizrachi Ilene.(2020). GenBank. Nucleic Acids Res, 48(null), D84-D86.
doi:10.1093/nar/gkz956.
44. Koren Sergey., Walenz Brian P., Berlin Konstantin., Miller Jason R., Bergman Nicholas H., Phillippy Adam M.(2017). kCanu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. Genome Res, 27(5), 722-736.
doi:10.1101/gr.215087.116.
45. Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler,QUAST: quality assessment tool for genome assemblies,Bioinformatics (2013) 29 (8): 1072-1075.doi: 10.1093/bioinformatics/btt086.
46. Parks Donovan H., Imelfort Michael., Skennerton Connor T., Hugenholtz Philip., Tyson Gene W.(2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res, 25(7), 1043-55.
doi:10.1101/gr.186072.114.
47. Manni Mosè., Berkeley Matthew R., Seppey Mathieu., Simão Felipe A., Zdobnov Evgeny M.(2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol, undefined(undefined), undefined.
doi:10.1093/molbev/msab199.

48. Seemann Torsten.(2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-9. doi:10.1093/bioinformatics/btu153.
49. Chaudhari, Narendrakumar M., Vinod Kumar Gupta, and Chitra Dutta. "BPGA:an ultra-fast pan-genome analysis pipeline." Scientific reports 6, 24373 (2016). doi: 10.1038/srep24373.
50. Huerta-Cepas Jaime., Szklarczyk Damian., Heller Davide., Hernández-Plaza Ana., Forslund Sofia K., Cook Helen., Mende Daniel R., Letunic Ivica., Rattei Thomas., Jensen Lars J., von Mering Christian., Bork Peer.(2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res, 47(null), D309-D314. doi:10.1093/nar/gky1085.
51. Carlos P. Cantalapiedra, Ana Hernandez-Plaza, Ivica Letunic, Peer Bork, and Jaime Huerta-Cepas. (2021).eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. bioRxiv. doi:10.1101/2021.06.03.446934.
52. Gardner Shea N., Slezak Tom., Hall Barry G.(2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics, 31(17), 2877-8. doi:10.1093/bioinformatics/btv271.
53. Stamatakis Alexandros.(2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-3. doi:10.1093/bioinformatics/btu033.

54. Letunic Ivica., Bork Peer.(2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*, 49(null), W293-W296. doi:10.1093/nar/gkab301.
55. Larsen Mette V., Cosentino Salvatore., Rasmussen Simon., Friis Carsten., Hasman Henrik., Marvig Rasmus Lykke., Jelsbak Lars., Sicheritz-Pontén Thomas., Ussery David W., Aarestrup Frank M., Lund Ole.(2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol*, 50(4), 1355-61. doi:10.1128/JCM.06094-11.
56. Zankari Ea., Hasman Henrik., Cosentino Salvatore., Vestergaard Martin., Rasmussen Simon., Lund Ole., Aarestrup Frank M., Larsen Mette Voldby.(2012). Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*, 67(11), 2640-4. doi:10.1093/jac/dks261.
57. Alcock Brian P., Raphenya Amogelang R., Lau Tammy T Y., Tsang Kara K., Bouchard Mégane., Edalatmand Arman., Huynh William., Nguyen Anna-Lisa V., Cheng Annie A., Liu Sihan., Min Sally Y., Miroshnichenko Anatoly., Tran Hiu-Ki., Werfalli Rafik E., Nasir Jalees A., Oloni Martins., Speicher David J., Florescu Alexandra., Singh Bhavya., Faltyn Mateusz., Hernandez-Koutoucheva Anastasia., Sharma Arjun N., Bordeleau Emily., Pawlowski Andrew C., Zubyk Haley L., Dooley Damion., Griffiths Emma., Maguire Finlay., Winsor Geoff L., Beiko Robert G., Brinkman Fiona S L., Hsiao William W L., Domselaar Gary V., McArthur Andrew G.(2020). CARD 2020: antibiotic

resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res*, 48(null), D517-D525. doi:10.1093/nar/gkz935.

58. Liu Bo., Zheng Dandan., Jin Qi., Chen Lihong., Yang Jian.(2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*, 47(null), D687-D692. doi:10.1093/nar/gky1080.

59. Nourdin-Galindo Guillermo., Sánchez Patricio., Molina Cristian F., Espinoza-Rojas Daniela A., Oliver Cristian., Ruiz Pamela., Vargas-Chacoff Luis., Cárcamo Juan G., Figueroa Jaime E., Mancilla Marcos., Maracaja-Coutinho Vinicius., Yañez Alejandro J.(2017). *Piscirickettsia salmonis* Comparative Pan-Genome Analysis of Reveals Genomic Divergences within Genogroups. *Front Cell Infect Microbiol*, 7(undefined), 459. doi:10.3389/fcimb.2017.00459.

60. Sievers Fabian., Wilm Andreas., Dineen David., Gibson Toby J., Karplus Kevin., Li Weizhong., Lopez Rodrigo., McWilliam Hamish., Remmert Michael., Söding Johannes., Thompson Julie D., Higgins Desmond G.(2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7(undefined), 539. doi:10.1038/msb.2011.75.

61. Robert Xavier., Gouet Patrice.(2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, 42(null), W320-4. doi:10.1093/nar/gku316.

62. Krzywinski Martin., Schein Jacqueline., Birol Inanç., Connors Joseph., Gascoyne Randy., Horsman Doug., Jones Steven J., Marra Marco A.(2009). Circos: an information aesthetic for comparative genomics. *Genome Res*, 19(9), 1639-45.
doi:10.1101/gr.092759.109.
63. Tettelin Hervé., Massignani Vega., Cieslewicz Michael J., Donati Claudio., Medini Duccio., Ward Naomi L., Angiuoli Samuel V., Crabtree Jonathan., Jones Amanda L., Durkin A Scott., Deboy Robert T., Davidsen Tanja M., Mora Marirosa., Scarselli Maria., Margarit y Ros Immaculada., Peterson Jeremy D., Hauser Christopher R., Sundaram Jaideep P., Nelson William C., Madupu Ramana., Brinkac Lauren M., Dodson Robert J., Rosovitz Mary J., Sullivan Steven A., Daugherty Sean C., Haft Daniel H., Selengut Jeremy., Gwinn Michelle L., Zhou Liwei., Zafar Nikhat., Khouri Hoda., Radune Diana., Dimitrov George., Watkins Kisha., O'Connor Kevin J B., Smith Shannon., Utterback Teresa R., White Owen., Rubens Craig E., Grandi Guido., Madoff Lawrence C., Kasper Dennis L., Telford John L., Wessels Michael R., Rappuoli Rino., Fraser Claire M.(2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A*, 102(39), 13950-5. doi:10.1073/pnas.0506758102.
64. Diene Seydina M., Merhej Vicky., Henry Mireille., El Filali Adil., Roux Véronique., Robert Catherine., Azza Saïd., Gavory Frederick., Barbe Valérie., La Scola Bernard., Raoult Didier., Rolain Jean-Marc.(2013). The rhizome of the multidrug-resistant

Enterobacter aerogenes genome reveals how new "killer bugs" are created because of a
sympatric lifestyle. Mol Biol Evol, 30(2), 369-83. doi:10.1093/molbev/mss236.

65. Galperin Michael Y., Wolf Yuri I., Makarova Kira S., Vera Alvarez Roberto.,
Landsman David., Koonin Eugene V.(2021). COG database update: focus on microbial
diversity, model organisms, and widespread pathogens. Nucleic Acids Res, 49(null),
D274-D281. doi:10.1093/nar/gkaa1018.

66. Jackson Robert W., Vinatzer Boris., Arnold Dawn L., Dorus Steve., Murillo
Jesús.(2011). The influence of the accessory genome on bacterial pathogen evolution.
Mob Genet Elements, 1(1), 55-65. doi:10.4161/mge.1.1.16432

67. McManus Brenda A., Coleman David C., Deasy Emily C., Brennan Gráinne I., O'
Connell Brian., Monecke Stefan., Ehricht Ralf., Leggett Bernadette., Leonard Nola.,
Shore Anna C.(2015). Comparative Genotypes, Staphylococcal Cassette Chromosome
mec (SCCmec) Genes and Antimicrobial Resistance amongst Staphylococcus
epidermidis and Staphylococcus haemolyticus Isolates from Infections in Humans and
Companion Animals. PLoS One, 10(9), e0138079. doi:10.1371/journal.pone.0138079.

68. Sun Z, Zhou D, Zhang X, Li Q, Lin H, Lu W, Liu H, Lu J, Lin X, Li K, Xu T, Bao Q
and Zhang H (2020) Determining the Genetic Characteristics of Resistance and Virulence
of the “Epidermidis Cluster Group” Through Pan-Genome Analysis. Front. Cell. Infect.
Microbiol. 10:274. doi: 10.3389/fcimb.2020.00274.

69. Oh, J., Byrd, A. L., Park, M., Program, N. C. S., Kong, H. H., and Segre, J. A. (2016). Temporal stability of the human skin microbiome. *Cell* 165, 854 – 866. doi: 10.1016/j.cell.2016.04.008.
70. Ziebuhr, W., Hennig, S., Eckart, M., Kranzler, H., Batzilla, C., and Kozitskaya, S. (2006). Nosocomial infections by *Staphylococcus epidermidis*: how a commensal bacterium turns into a pathogen. *Int. J. Antimicrob. Agents* 28(Suppl. 1), S14 – S20. doi: 10.1016/j.ijantimicag.2006.05.012.
71. Kleinschmidt, S., Huygens, F., Faoagali, J., Rathnayake, I. U., and Hafner, L. M. (2015). *Staphylococcus epidermidis* as a cause of bacteremia. *Future Microbiol.* 10, 1859 – 1879. doi: 10.2217/fmb.15.98.
72. Durand M L.(2013). Endophthalmitis. *Clin Microbiol Infect*, 19(3), 227-34. doi:10.1111/1469-0691.12118.
73. Georgiades, K., and Raoult, D. (2010). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151.
74. Zhang Yue., Liu Zhi-Rong., Chen Hui., Fan Ying-Chuan., Duo Ji., Zheng Hong., Wang Guang-Jin., Li Yu-Chan., Jiachu Dan-Ba., Zewang Ge-Ma.(2013). Comparison on conjunctival sac bacterial flora of the seniors with dry eye in Ganzi autonomous prefecture. *Int J Ophthalmol*, 6(4), 452-7. doi:10.3980/j.issn.2222-3959.2013.04.08.
75. Ettema Thijs J G., Brinkman Arie B., Tani Travis H., Rafferty John B., Van Der Oost John.(2002). A novel ligand-binding domain involved in regulation of amino acid

metabolism in prokaryotes. *J Biol Chem*, 277(40), 37464-8.

doi:10.1074/jbc.M206063200.

76. Henikoff S., Haughn G W., Calvo J M., Wallace J C.(1988). A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A*, 85(18), 6602-6.

doi:10.1073/pnas.85.18.6602.

77. Cao Zhenping., Casabona M Guillermina., Kneuper Holger., Chalmers James D., Palmer Tracy.(2016). The type VII secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria. *Nat Microbiol*, 2(defined), 16183.

doi:10.1038/nmicrobiol.2016.183.

78. Park Joo Youn., Kim Jong Wan., Moon Bo Youn., Lee Juyeun., Fortin Ye Ji., Austin Frank W., Yang Soo-Jin., Seo Keun Seok.(2015). Characterization of a novel two-component regulatory system, HptRS, the regulator for the hexose phosphate transport system in *Staphylococcus aureus*. *Infect Immun*, 83(4), 1620-8.

doi:10.1128/IAI.03109-14.

79. Vazquez Vanessa., Liang Xiaowen., Horndahl Jenny K., Ganesh Vannakambadi K., Smeds Emanuel., Foster Timothy J., Hook Magnus.(2011). Fibrinogen is a ligand for the *Staphylococcus aureus* microbial surface components recognizing adhesive matrix molecules (MSCRAMM) bone sialoprotein-binding protein (Bbp). *J Biol Chem*, 286(34), 29797-805. doi:10.1074/jbc.M110.214981.

80. Qin, L., Mccausland, J. W., Cheung, G. Y., and Otto, M. (2016). PSM-Mec-A virulence determinant that connects transcriptional regulation, virulence, and antibiotic resistance in *Staphylococci*. *Front. Microbiol.* 7:1293. doi: 10.3389/fmicb.2016.01293.

Tables

Table1 Genomic characteristics of 11 ocular trauma-source *Staphylococcus epidermidis*

	Mapping Reads	Unmappin -g Reads	Mean Coverage Depth	Coverage Depth<10 (%)	GC content	Contig number	Genome size (Mb)	CDS	tRNA	rRNA
O106	1063834	0	4524	<0.001%	32.27%	2	2.54	2360	60	19
O107	1635910	0	6830	0	32.15%	3	2.56	2355	60	19
O108	1148008	0	5347	0	32.24%	1	2.49	2247	60	22
O109	1457938	0	6404	0.0023%	32.10%	3	2.56	2391	59	19
O110	1523429	0	6511	0	32.05%	3	2.67	2446	60	19
O112	1629802	0	6936	0	32.17%	2	2.54	2313	59	19
O113	1495273	0	6420	0.016%	32.04%	4	2.62	2455	60	19
O114	1485509	0	6590	<0.001%	32.15%	4	2.58	2345	60	19
O115	988780	0	4438	0	32.15%	2	2.51	2272	60	19
O116	1110019	0	4641	0.1%	32.01%	6	2.79	2692	60	19
O117	1130833	0	5414	<0.001	32.27%	1	2.48	2265	61	19

Table2 *Mec* gene carried rate and antibiotic susceptibility testing results of 11 ocular trauma-source *Staphylococcus epidermidis*

	O106	O107	O108	O109	O110	O112	O113	O114	O115	O116	O117
<i>mecA</i>	+	-	+	-	+	+	-	-	-	-	+
<i>mecRI</i>	+	-	-	-	-	-	-	-	-	-	+
oxacillin	R	S	R	S	R	R	S	S	S	S	R
cefoxitin	R	S	R	S	R	R	S	S	S	S	R

Figures

FIGURE | 1 Circular genome maps of representative ocular strain. From the outer to the inner circle: (1) scale marks of genomes; (2) assigned COG classes of protein-coding genes (CDSs) on the forward strand; (3) reverse strand CDSs; (4) tRNA (blue) and rRNA (red) genes on the forward strand; (5) tRNA (blue) and rRNA (red) genes on the reversed strand; (6) GC content (swell fire red/sky blue indicates higher/lower G + C compared with the average G + C content); (7) GC skew (black/orange indicate positive/negative values).

829

830

831

FIGURE | 2 Pan-genome and phylogenetic feature of *S. epidermidis* from different host niches and health state. A. Gene accumulation curves for pangenome (red) and core-genome (blue) of collected all strains. B. Gene accumulation curves for pangenome (red) and core-genome (blue) of ocular sources strains. C. COG functional categories from the pan-genomes within strains from different niches. Involved COG categories are as follows: [B]Chromatin structure & dynamics; [C]Energy production & conversion; [D]Cell cycle control, cell division, chromosome partitioning; [E]Amino acid transport & metabolism; [F]Nucleotide transport & metabolism; [G]Carbohydrate transport & metabolism; [H]Coenzyme transport & metabolism; [I]Lipid transport & metabolism; [J]Translation, ribosomal structure & biogenesis; [K]Transcription; [L]Replication, recombination & repair; [M]Cell wall/membrane/envelope biogenesis; [N]Cell motility; [O]Post-translational modification, protein turnover & chaperones; [P]Inorganic ion transport & metabolism; [Q]Secondary metabolites biosynthesis, transport & catabolism; [T]Signal transduction mechanisms; [U]Intracellular trafficking, secretion & vesicular transport; [V]Defense mechanisms; [W]Extracellular structures. Poorly characterized COG categories contains [R]General function prediction only and [S]Function unknown. D. Phylogenetic core SNP maximum likelihood tree was constructed for 187 genomes. The blue branch represent strains from diseased hosts while healthy source strains put color on red. Different color of strains ID stand for different host niches: blood (blue), ocular (green), skins (yellow) and respiratory tract (brick-red).

Legends on the left stand for colors of sequence type(ST) from multi-locus sequence typing.

FIGURE | 3 Heatmap of virulence genes in all 187 *S. epidermidis* strains. The red square indicates the presence of the gene, while the white square indicates absence. Legends on the right stand for colors of different host healthy state, isolates source and virulence factor categories.

FIGURE | 4 Gene function difference between healthy and disease source genomes. A. Heatmap of enriched genes (fisher's exact test, adjusted $P < 0.05$) between diseased and healthy sources strains. Only the accessory genes were shown. Gene clusters present in all genomes (core gene) or present in only a single genome (unique genes) are omitted. The red color stands for genes that existed and the blue color for missing ones. B. Pie graph of KEGG function related with enriched genes from different health source. Bigger pie represented KEGG function proportion of 45 enriched genes showed in figure 3A, while three smaller pie were account for function categories of metabolism, genetic information process, human diseases, signaling and cellular processes, respectively.

876 genes and disease versus health source; P values were calculated using Fisher's exact test.

877 C. Multiple alignment of virulence gene sdrG and non virulence gene sdrG protein

878 sequences across the collected blood source *S.epidermidis* (blue), intraocular isolates

879 (green), *staphylococcus aureus* and sdrG positive skin source strains.

FIGURE | 6 Presence of antibiotic resistant genes across *Staphylococcus epidermidis*

strains. Heatmap indicates of 41 antibiotic resistant genes involved. The red color stands

for genes that existed and the dark blue color for missing ones.