A peer-reviewed version of this preprint was published in PeerJ on 20 November 2019.

View the peer-reviewed version (peerj.com/articles/8068), which is the preferred citable publication unless you specifically need to cite this preprint.

Cabrera-Contreras R, Santamaría RI, Bustos P, Martínez-Flores I, Meléndez-Herrada E, Morelos-Ramírez R, Barbosa-Amezcua M, González-Covarrubias V, Silva-Herzog E, Soberón X, González V. 2019. Genomic diversity of prevalent *Staphylococcus epidermidis* multidrug-resistant strains isolated from a Children's Hospital in México City in an eight-years survey. PeerJ 7:e8068 https://doi.org/10.7717/peerj.8068

Genomic analysis of prevalent *Staphylococcus epidermidis* multi-drug resistant strains isolated during eight years in a single children hospital in México City.

Roberto Cabrera-Contreras [Corresp., 1], **Rosa I Santamaría** [2], **Patricia Bustos** [2], **Irma Martínez-Flores** [2], **Enrique Meléndez** [1], **Rubén Morelos** [1], **Martín Barbosa-Amezcua** [3], **Vanessa González-Covarrubias** [3], **Eugenia Silva-Herzog** [3], **Xavier Soberón** [3], **Víctor González** [Corresp. 2]

[1] Laboratorio de Patogenicidad Bacteriana, Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México

[2] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

[3] Instituto Nacional de Medicina Genómica, Mexico City, México

Corresponding Authors: Roberto Cabrera-Contreras, Víctor González
Email address: cabreracontrerasr@yahoo.com, vgonzal@ccg.unam.mx

Staphylococcus epidermidis is a human commensal and pathogen worldwide distributed. In this work, we surveyed for multi-resistant *S. epidermidis* strains in eight years at a children health-care unit in México City. Multidrug-resistant *S. epidermidis* were present in all years of the study. Resistance to methicillin, beta-lactams, fluoroquinolones, and macrolides were included. To understand the genetic basis of antibiotic resistance and its association with virulence and gene exchange, we sequenced the genomes of 17 *S. epidermidis* isolates. Whole-genome nucleotide identities between all the pairs of *S. epidermidis*strains were about 97% to 99%. We inferred a clonal structure and eight Multilocus Sequence Types (MLST´s) in the *S. epidermidis* sequenced collection. The profile of virulence includes genes involved in biofilm formation and phenol-soluble modulins (PSMs). However, half of the *S. epidermidis* analyzed lacked the *ica*operon for biofilm formation. Likely, they are commensal *S. epidermidis* strains but multi-antibiotic resistant. Uneven distribution of insertion sequences, phages, and CRISPR-Cas immunity phage systems suggest frequent horizontal gene transfer. Rates of recombination between *S. epidermidis* strains were more prevalent than the mutation rate and affected the whole genome. Therefore, the multidrug resistance, independently of the pathogenic traits, might explain the persistence of specific highly adapted *S. epidermidis* clonal lineages in nosocomial settings.

Genomic analysis of prevalent *Staphylococcus epidermidis* antibiotic multi-resistant isolated during eight years in a single children hospital in México City

Roberto Cabrera-Contreras[1], Rosa I. Santamaría[2], Patricia Bustos[2], Irma Martínez-Flores[2], Enrique Meléndez-Herrada[1], Rubén Morelos-Ramírez[1], Martín Barbosa-Amezcua[3], Vanessa González[3], Eugenia Silva-Herzog[3], Xavier Soberón[3] and Víctor González[2].

[1] Laboratorio de Patogenicidad Bacteriana del Departamento de Salud Pública de la Facultad de Medicina, Universidad Nacional Autónoma de México. 04510. Ciudad de México. México.
[2] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad s/n Col. Chamilpa, Cuernavaca, Morelos CP 62210. México.
[3] Instituto Nacional de Medicina Genómica, Periférico Sur No. 4809, Col. Arenal Tepepan, Delegación Tlalpan, Ciudad de México, C.P. 14610, México.

Corresponding authors:
Roberto Cabrera-Contreras
Av. Universidad 3000, Coyoacán, Ciudad de México. CP 04510. México.
E-mail: cabreracontrerasr@yahoo.com
Xavier Soberón
Periférico Sur 4809, Arenal Tepepan, Ciudad de México. 14610. México.
E-mail: xsoberon@inmegen.gob.mx
Víctor González
Av. Universidad N/C, Chamilpa, Morelos. 62219. México
E-mail: vgonzal@ccg.unam.mx

Abstract

Staphylococcus epidermidis is a human commensal and pathogen worldwide distributed. In this work, we surveyed for multi-resistant *S. epidermidis* strains in eight years at a children health-care unit in México City. Multidrug-resistant *S. epidermidis* were present in all years of the study. Resistance to methicillin, beta-lactams, fluoroquinolones, and macrolides were included. To understand the genetic basis of antibiotic resistance and its association with virulence and gene exchange, we sequenced the genomes of 17 *S. epidermidis* isolates. Whole-genome nucleotide identities between all the pairs of *S. epidermidis* strains were about 97% to 99%. We inferred a clonal structure and eight Multilocus Sequence Types (MLST´s) in the *S. epidermidis* sequenced collection. The profile of virulence includes genes involved in biofilm formation and phenol-soluble modulins (PSMs). However, half of the *S. epidermidis* analyzed lacked the *ica* operon for biofilm formation. Likely, they are commensal *S. epidermidis* strains but multi-antibiotic resistant. Uneven distribution of insertion sequences, phages, and CRISPR-Cas immunity phage systems suggest frequent horizontal gene transfer. Rates of recombination between *S. epidermidis* strains were more prevalent than the mutation rate and affected the whole genome. Therefore, the multidrug resistance, independently of the pathogenic traits, might explain the persistence of specific highly adapted *S. epidermidis* clonal lineages in nosocomial settings.

Introduction

Staphylococcus epidermidis (*SE*) is a typical commensal bacterium of the human skin microbiome (Byrd et al., 2018). However, some *SE* strains behave as pathogens colonizing surgery wounds, medical devices, and in some circumstances, they reach the human bloodstream causing severe bacteremia and potential mortality (Otto, 2009; Chessa et al., 2015). Children are especially prone to acquire methicillin-resistant *SE* strains in perinatal hospitals (Villari et al., 2000; Marchant et al., 2013). The genetic basis for pathogenicity of *SE* strains includes genes related to biofilm formation (adhesion), phenol soluble modulins (PSMs), and diverse Mobile Genetic Elements (MGEs) like phages, insertion sequences (ISs), and pathogenicity islands, that may be associated with the transfer of antibiotic and virulence traits (Miragaia et al., 2009; Bouchami et al., 2016; Rolo et al., 2017) (Kozitskaya et al., 2005; Conlan et al., 2012; Meric et al., 2015). There is no clear genetic distinction between pathogenic and commensal non-pathogenic *SE* strains, even though nosocomial strains are enriched in virulence and antibiotic resistance genes (Kozitskaya et al., 2005; Conlan et al., 2012; Meric et al., 2018). It has been proposed that these genes are within the pool of accessory genome, mobilized within and between species (Meric et al., 2015; Rolo et al., 2017).

The population structure of *SE* is essentially clonal as determined by Multilocus Sequence Typing (MLST) (Miragaia et al., 2002; Miragaia et al., 2007; Thomas et al., 2007). Although, nosocomial *SE* strains show high genetic diversity among isolates from distant geographic locations (Miragaia et al., 2007) several clonal complexes are disseminated globally including the ST2, ST5, and ST23 which indeed are the most frequently found in clinical environments (Miragaia et al., 2007; Lee et al., 2018). Besides the clonal structure, recombination has been estimated to occur two times more frequently than mutation (Miragaia et al., 2007). Genomic

analysis has calculated that about 40% of the core genes of *SE* had undergone recombination (Meric et al., 2015). It suggests that recombination might be a general property of *SE* for rapid evolution in clinical settings but conserving high linkage disequilibrium between alleles. In this case, recombination might act as a cohesive force, maintaining the clonal population structure but allowing clones to diverge.

Multidrug resistance is a major concern in most health care settings worldwide because of the difficulties associated with clinical treatment and the potential of dissemination. It has been reported that the prevalent *SE* clonal lineages ST5, ST12, and ST23 often display high resistance toward most of the antibiotic classes (Widerstrom et al., 2012; Martinez-Melendez et al., 2016). Furthermore, there is evidence of increasing of rifampicin resistance over time in clinical *SE* isolates that belong to ST2 and ST23 from the USA, Australia, and some European countries (Lee et al., 2018). Mutations in the *rpoB* gene conferring rifampicin resistance have emerged independently, suggesting the view that a limited number, well adapted multi-resistant clonal *SE* lineages, prevail in clinical settings (Martinez-Melendez et al., 2016; Lee et al., 2018). Reports of the profiles of antibiotic multi-resistance changes over time are scarce (Lee et al., 2018). Surveillance of rifampicin resistance in a study involving 24 countries and 96 institutions suggests annual local variations for this antibiotic (Lee et al., 2018). However, the correct identification of pathogenic *SE* strains and its drug resistance profile will contribute to prevent and treat these bacterial infections in the clinic.

In this work, we evaluated the prevalence of *SE* in comparison with other staphylococci isolated in a single hospital in México City during eight years period. Then, we aimed to assess the antibiotic resistance profiles of *SE* isolates, and through genomics, to know the genome structure of a selected set of antibiotic multi-resistant *SE* INPer strains. In this context, we provide a genome-wide measure of recombination, to shed light on the mechanisms of *SE* genetic diversification and the evolution of multi-resistance in local sites.

Materials & Methods

Phenotypic analysis. This study was carried out following the recommendations of the ethics review committee of the Facultad de Medicina de la UNAM. The *Staphylococcus epidermidis* strains used in this work were obtained by the donation from the microbiology collection of the Instituto Nacional de Perinatología “Isidro Espinosa de los Reyes” (INPer). A consent form was not required. Original identification keys and clinical data concerning the isolates are maintained under control of INPer. INPer strain identifiers were substituted by common names to prevent further interpretations of the data. Authors do not have access in any form to the specific clinical information of strains and patients.

Staphylococci strains were isolated at the National Institute of Perinatology (Instituto Nacional de Perinatología “INPer” “Dr. Isidro Espinosa de los Reyes” SSA) under the ethical standards of the Medicine Faculty of the National Autonomous University of Mexico. Conventional phenotypic analyses, including coagulase test, were performed on the 17 *SE* strains. The biotype and antibiotype for 17 antibiotics were determined using the automatic VITEK equipment. The antimicrobials tested were as follows: Amoxicillin/Clavulanic Acid, Ampicillin, Cefazolin, Ciprofloxacin, Clindamycin, Erythromycin, Gentamicin, Imipenem, Levofloxacin, Linezolid. Oxacillin, Penicillin, Rifampicin, Synercid, Tetracycline, Trimethoprim/Sulfamethoxazole, and Vancomycin.

Genomic sequencing and annotation. Genomic DNA was isolated by using Lysostaphin-Lysozyme and isopropanol precipitation (Coll, 2005). Nextera libraries were generated from genomic DNA and sequenced using a paired-end base dual index, performed on a MySeq Illumina unit producing 1-2 millions of reads pairs per library by ~80x genome coverage reads (Andrews, 2010) (Bolger et al., 2014). Assemblies were made with Spades 3.6.0 in genomes sequenced in high coverage (50x to 100x). Genome annotations were obtained from the PATRIC server (Wattam et al., 2018) (https://www.patricbrc.org/), through the automated bioinformatic method RAST (Rapid Annotation Subsystem Technologies) (Antonopoulos et al., 2017). Gene annotation of antibiotic resistance and virulence-related genes were obtained from the section of Special Genes of PATRIC. Annotations for virulome were supported with other sources of information such as PATRIC VF, VFDB, and Victors, the information obtained refers to genes identified or those associated with virulence proposed in the literature. For the prediction of antibiotic resistance genes, the PATRIC server was used to search specialty genes and is supported for this search with the external databases CARD (Comprehensive Antibiotic Resistance Database) and NDARO (National Database of Antibiotic Resistant Organism) (Davis et al., 2016).

Genome comparisons and pangenome modeling. The Average Nucleotide Identity by MUMmer (ANIm) and the Genomic Coverage (G_{cov}) were calculated with the JSspecies program, with MUMmer used as a pairwise comparison tool for pairs of *SE* INPer genomes (Richter et al., 2016). The pangenome was modeled using the GET_HOMOLOGUES package and its defaults values (Contreras-Moreira and Vinuesa, 2013). Briefly, groups of orthologues proteins were computed using the Ortho-MCL integrated into GET_HOMOLOGUES. Accessory genes (unique genes and genes present in at least in two genomes but not in the all genomes), were obtained from the cloud, shell, and soft-core pangenome components according to GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013). The distribution of accessory genes in *SE* genomes was performed with the heatmap.2 function of the R´s ggplot2 package.

Phylogeny. A consensus core genome calculated from the pangenome model of 29 *SE* genomes was obtained with GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013). The resulting 1575 core protein clusters were subject to multiple alignments with MUSCLE (Edgar, 2004), gaps removed with TrimAl v2.1, and were concatenated using homemade Perl scripts (Capella-Gutierrez and Gabaldon, 2013). Phylogenetic trees were constructed by the maximum likelihood (ML) method based on the substitution matrix of JTT (Jones, Taylor, Thornton), with 1000 bootstrap replicates, using RaxML program (Stamatakis, 2015). To draw and edit the phylogenetic trees, we used the iTool program (Letunic and Bork, 2019).

Recombination. Inference of recombination was performed with ClonalFrameML (Didelot and Wilson, 2015). First, we ran GET-HOMOLOGUES to obtain the common protein clusters encoded by the genomes of each set of *SE* strains (Contreras-Moreira and Vinuesa, 2013). Second, they were converted to nucleotide sequences and concatenated using homemade Perl scripts. Multiple alignments were made as described above for phylogeny construction. Third, a RAxML (Stamatakis, 2015) tree was done to obtain the Newick format and the transition/transversion parameter for running ClonalFrameML under default parameters (Didelot and Wilson, 2015). Z-score statistics were obtained for all the sets of *SE* genomes and p-values using the web application Z Score Calculator

(https://www.socscistatistics.com/tests/ztest/zscorecalculator.aspx). BoxPlots were performed with the R ggplot2 system.

Mobile elements identification. Prophages were identified with PHAST (Arndt et al., 2017). Only predictions ranked as "intact prophages" were considered for analysis. IS, CRISPR-Cas elements and spacer sequences were obtained from PATRIC annotations (Antonopoulos et al., 2017). Then, IS were classified into families by BLASTx comparison with the ISfinder (Siguier et al., 2006). To determine the identity of the DNA sequences of the spacers located in the CRISPR-Cas elements, they were compared by BLASTn with the NCBI virus database. Only identical matches with a phage sequence in the database were recorded (Paez-Espino et al., 2016).

GenBank accession numbers. *S. epidermidis* of the INPer collection used in this work were uploaded in GenBank with the following Biosample identifiers: SAMN11086744, SAMN11086745, SAMN11086746, SAMN11086747, SAMN11086748, SAMN11086749, SAMN11086750, SAMN11086751, SAMN11086752, SAMN11086753, SAMN11086754, SAMN11086755, SAMN11086756, SAMN11086757, SAMN11086758, SAMN11086759, SAMN11086760. The accession numbers for the genomes of reference *S. epidermidis* strains are listed in Table S3.

Results

Survey of antibiotic multi-resistant intrahospital staphylococci in eight years period. We studied the incidence of staphylococci species and strains multi-resistant to antibiotics in the above-mentioned children hospital from 2006 to 2013. A total of 822 staphylococci strains were recovered from distinct infection sites of newborns and adults. *Staphylococcus* species were identified by standard clinical methods, including tests for biofilm formation, coagulase reaction, and resistance to 17 antibiotics (see methods). The analysis showed that *SE* strains were the most abundant (595 strains), followed by *S. aureus* (146 strains), and other coagulase-negative *Staphylococcus* (CoNS) species in low proportion (81 strains) (Fig. S1). The *SE* group was constituted mainly by strains recovered from blood samples and catheters (Fig. S1).

The number of antibiotic multi-resistant *SE* strains per year showed similar profiles along the studied period (Fig. 1, A-H). *SE* strains multi-resistant to nine antibiotics were the most frequently found throughout the eight years. In contrast, *S. aureus* strains showed resistance to about four antibiotics as the most common profile (Fig. S2). Similarly, the resistance to each antibiotic remained without change, along with the eight-year study (Fig. 1, I-P). These results suggest the persistence of a multidrug-resistant clone or few clones of *SE* at the hospital during the eight years, or frequent horizontal exchange of genes encoding antibiotic resistance between *SE* clones. To study these alternatives, we choose one or more *SE* strains per year to totalize 17 clinical *SE* INPer coagulase-negative (CoNS) strains (Fig. 1). These particular strains came from nosocomial infections of 14 newborns and three adults, isolated from three different infection sites: blood (8 strains), catheters (7 strains), cerebrospinal fluid (1 strain), and soft tissue (1 strain) (Table 1).

Broad gene catalog from draft genomes. To understand the genetic basis of antibiotic resistance and pathogenicity of the nosocomial *SE* INPer strains, we sequenced the whole

genome sequence of the 17 strains. After testing different parameters with the assembler programs Velvet and Spades (Zerbino, 2010; Bankevich et al., 2012), we obtained draft genomes assemblies consisting about 92 up to 432 contigs with a 60-70x average sequence coverture per genome (Table S1). To assert that the 17 assemblies represent a substantial part of the *SE* genomes, we compared the total genome length, and the number and length of the predicted ORFs, with 11 complete genomes of *SE* downloaded from GenBank (Table S2). There were no differences between the genome length of the *SE* INPer genomes and the complete genomes from the GenBank (unpaired t-test = 2.33; p-value = 0.022), in ORFs number (unpaired t-test = -1.68; p-value = 0.117), or ORFs length (unpaired t-test = 2.60; p-value = 0.014) (Table S2). Therefore, we can conclude that the *SE* INPer genome sequences obtained here provide a broad catalog of genes per genome useful for comparative genomics.

Genomic, pangenomic, and phylogenetic relationships among *SE* INPer isolates. To define the genomic similarity between the 17 clinical *SE* INPer strains, we performed total pairwise nucleotide identity estimates (Average Nucleotide Identity by Mummer, ANIm) (Richter et al., 2016). The *SE* INPer, showed high genomic ANIm values of about 99%, covering more than 90% of the genome length (Fig. S3). One exception is strain S10, which showed ANIm values of about 97% concerning the rest of the strains.

Besides the genomic identity between *SE* INPer isolates, we wanted to investigate the extent of their genetic variability by performing a pangenome model. To this end, the *SE* INPer collection was complemented by the inclusion of 12 complete genomes of *SE* strains downloaded from GenBank (Table S3, NCBI complete genomes). The model obtained using the GET_HOMOLOGUES software package (Contreras-Moreira and Vinuesa, 2013), indicated an open pangenome (Fig. S4). The core genome component for the 29 *SE* strains was predicted to consist of about 1575 gene clusters, whereas the sum of genes unevenly distributed in the 29 *SE* genomes (accessory component) contains 4360 gene clusters.

To know the phylogenetic relationship of the *SE* INPer strains in the context of reference *SE* strains, we did an un-rooted ML phylogenetic tree using the predicted 1575 concatenated core proteins (Fig. 2). There were three clades separated by the largest branches in the tree that comprise most of the *SE* INPer strains, and one or more *SE* strains isolate worldwide (Fig. 2, clades B, C, D). The clade marked as D in the tree consisted of two groups, one of which includes only reference *SE* strains, whereas the other had most of the *SE* INPer strains. Strain S10, the most different strain by ANIm in the *SE* INPer collection, was grouped in the clade A with the commensal strains *SE* ATCC12228 (2) and *SE* 14.1.R1 isolated from USA and Denmark.

The distribution of accessory genes in individual *SE* strains coincided with the phylogenetic clades. The *SE* INPer strains grouped in three phylogenetic clades are also related by their similitude in the gene presence/absence profile (Fig. 3). Despite the high identity of the *SE* strains, there is a still considerable individual variation that may account for adaptations to local milieu.

Clonal structure. To investigate to which clonal ST complex the *SE* INPer strains belongs, we looked for the seven proteins of the *S. epidermidis* MLST scheme, and compared them with their

respective alleles in the *Staphylococcus epidermidis* MLST database (https://pubmlst.org/sepidermidis/; Table 1; see methods) (Feil et al., 2004; Thomas et al., 2007). The analysis showed a total of 8 different STs; seven of them already recorded in the database. The S10 strain had an unassigned ST in the database and only differed by a single amino acid substitution in the YqiL protein (370L to C). The ST2, ST5, and ST23 are worldwide distributed and are the most represented in our sample (Miragaia et al., 2007; Lee et al., 2018). In agreement with global data, ST35, ST59, ST81, ST89 are less frequently represented. The clonal relationships among STs determined by eBURST, indicate that founder clones are ST2 and ST5, whereas the other four STs (ST 59, 81, 89 and 23), are peripheral clones mostly related to each than to the primary founder clones (Fig. S5).

Virulence genes. There are a few virulence genes characterized exclusively in *SE* in comparison to those found in *S. aureus* (Otto, 2009). Several known virulence genes of *SE* are shared by commensal and pathogenic strains (Otto, 2009). Among them, we found the cluster *icaADBCR* (biofilm formation) in nine out of the 17 *SE* INPer genomes analyzed, except for *icaC,* absent in strain S05 (Table 1). The phenol-soluble modulins (PSMs), involved in inflammatory response and lysis of leukocytes were present in the genomes of all *SE* INPer strains (Peschel and Otto, 2013). The novel ESAT 6 (*esaAB*, *essABC, and esxAB* genes*)* secretion system implicated in immune system evasion and neutrophil elimination was only present in S10 strain (Burts et al., 2005; Wang et al., 2016; Dai et al., 2017).

Antibiotic resistance genotype and phenotype. The *SE* INPer strains were tested for their susceptibility to methicillin and other β-lactams antibiotics. All the 17 *SE* strains have the β-lactamase gene (*blaZ*) and their regulators (*blaR* and *I*), which are probably responsible for the broad resistance spectrum to penicillin, carbapenems, and cephalosporins determined by VITEK system (Table 2). Only the INPer *SE* S10 and S16 strains were susceptible to methicillin while for the 14 resistant strains the presence of the *mecA* gene was confirmed in the genomes (Table 1). The gene *mecA* encodes for a penicillin-binding protein (PBP) carried in a mobile element known as *S*taphylococcal Chromosomal Cassette or SCC*mec* (International Working Group on the Classification of Staphylococcal Cassette Chromosome, 2009). By localizing the recombinases *ccrA*, *B*, and *C*, as well the *mecA* genes in the contigs of the respective genomes, and then constructing phylogenies including known SCC*mec* recombinase genes we classified the SCC*mec* types of the *SE* INPer strains (Table 1; Fig. S6). The analysis demonstrated the presence of the community-acquired SCC*mec* type IV cassette in 13 out of 14 methicillin-resistant strains. The *SE* INPer strains S10 and S16 lack the SCC*mec* cassette, and no *mecA* gene was detected. Although the S21 strain has a *mecA* gene, we were unable to find other gene elements to show the presence of a *mec* cassette. Moreover, INPer strains S07, S09, S13 strains contain an additional SCC*mec* type VIII cassette in tandem with the SCC*mec* IV cassette. The S07 strain carries a contig of 36535 bps of the SCC*mec* type IV and VIII, suggesting the probable structure of the recombined cassette (Fig. S7).

The genomic analysis revealed that some *SE* strains studied here included genes for resistance to fluoroquinolones, macrolides, sulfonamides, aminoglycosides, tetracycline, and other antibiotics not used as the first choice in clinical therapy (Table 2). The results given in Table 2 corroborate that in most of the cases, the probable gene responsible for the resistance is present in the genomes. Besides, non-synonymous mutations in the antibiotic target proteins GyrA and RpoB

were identified in some *SE* strains resistant to quinolones and rifampicin. Despite other INPer *SE* strains lack these mutations, they still were resistant to these antibiotics. Then, other mutations in the antibiotic target protein or other genetic mechanisms not yet known would be responsible for these resistances.

Mobile genetic elements. The genomic variability observed in the INPer *SE* strains suggests active processes of recombination and gene exchange. To study this concern, we first look for prophages and CRISPR-Cas related systems in the genomes. Prophage footprints were found in 10 out of 17 genomes of the *SE* strains. The most significant prophage hits detected by PHAST program, were for genomic regions spanning about 28 to 95 kb that include an attachment site, a signature of lysogenic phages (Arndt et al., 2017) (Table S4). In this analysis, prophages were found integrated into the genomes of some *SE* strains, such as CNPH82 found in the *SE* strains S14, S17, and S18 (Daniel et al., 2007). Some other prophages such as StB20 and SpBeta were present in S03 and S16 strains respectively and, the prophages IPLA5 in strain S07 and IPLA7 in S12 and S15 strains (Gutierrez et al., 2012). In the remaining *SE* strains, prophages sequences were not detected.

SE strains have also acquired defense mechanisms against phage infection. The search for CRISPR-Cas immune systems results in nine out of 17 *SE* INPer strains carrying a CRISPR-Cas Type III system. It is constituted by Cas1 and Cas2, responsible for spacer processing and insertion, ribonuclease Cas6, and the cascade proteins Csm1 to Csm6, involved in the processing of the target transcript (Table S5). The CRISPR-Cas Type III system has been already reported in *SE* to confer immunity to phages as well as to conjugative plasmids (Marraffini and Sontheimer, 2008; Marraffini, 2015). Although the enzymatic organization of the CRISPR-Cas systems is remarkably conserved in *SE*, there are variations in the array of repeats and spacers in the CRISPR loci. Three distinct types of identical repeated units of 30 or 36 nucleotides, associated with specific sequence spacers have been described (Marraffini, 2015). Three spacers that correspond to the CRISPR loci found in the strains S02, S05, and S24, match precisely with a sequence in the *Staphylococcus* phage PH15 genome for the first two strains and the *Staphylococcus* phage 6ec genome for the last (Daniel et al., 2007; Aswani et al., 2014).

SE strains harbor many ISs, belonging to different families (Table S6). The presence of IS256 has been found within pathogenic *SE,* associated with biofilm formation and virulence (Kozitskaya et al., 2004; Murugesan et al., 2018). Among the strains of our collection, there is a clear relationship between strains having the IS256 and the presence of the *ica* operon, confirming previous observations. Exceptionally, only the strain S21 has the *ica* genes but lacks IS256.

Recombination. The above results indicate that high frequency HGT and recombination might promote diversification of local *SE* populations. To evaluate this proposition, we measured the ratio of recombination to mutation (r/m), using the ClonalFrameML program (Didelot and Wilson, 2015) in 17 *SE* INPer genomes and several sets of *SE* genomes downloaded from GenBank (Fig. 2; Table S3). A median average r/m rate about 6.9 was calculated when the 17 *SE* INPer were tested, suggesting that nucleotide substitutions by recombination are more frequent than random point mutations (Vos and Didelot, 2009). Every COG class shows r/m values equal or higher than the estimate for the complete set of 17 *SE* genomes. Indeed, the r/m values on

virulence or antibiotic resistance gene class result similar to the other COGs involved in housekeeping functions.

To determine whether or not the recombination estimates were affected by the sample composition of *SE* strains, we design several control tests, with distinct groups of genomes. First, we discarded the most divergent S10 strain of the *SE* INPer collection and ran the ClonalFrameML test only with the 16 *SE* most related *SE* strains of the collection. As shown in Figure 4, the r/m rate for this set reduced to a median of four, and this value is not significantly different respect to the r/m of *SE* of the complete genome of *SE* strains obtained from the GenBank (z-score = 1.4, P-value 0.135) (Fig. 4, boxplot 25). Second, we computed the r/m rate in eight different sets randomly selected among 260 complete and draft genomes of *SE* strains from the GenBank (Fig. 4, 26- 33; Table S3). Some sets (Ctr2 and 8) display the lowest r/m values, whereas the rest control sets have r/m upper than two up to four. These results indicate that the strains sample composition influence the recombination estimates.

The RaxML nucleotide phylogenetic tree used as a reference to estimate recombination looks similar to the core protein phylogeny presented in Fig. 2; *SE* strains within clades maintain cohesive relationships (Fig. 5). However, multiple recombination events were detected in the ancestral nodes leading to the *SE* INPer strains. The most prominent branch (red dot line in Fig. 5) divided the *SE* strains into two large clades, one including the clade D and the other constituted by clade B and C. Within the different divergent lineages, recombination introduces much more nucleotide variants than mutation as presented before (r/m). These results indicate that a local hospital settings, *SE* strains may contain enough genomic diversity despite their close relationship with the main clonal ST complexes of worldwide distribution.

Discussion

SE is among the most common bacterial isolates found in the human skin microbiome (Byrd et al., 2018). It is also frequently recovered from bacteremia and sepsis samples in neonatal care clinic units, being its most probable etiological-agent (Otto, 2009; Byrd et al., 2018). In clinical practice, it is difficult to assess if *SE* clones are the causal agents of the disease or they are accidental, or opportunistic pathogens (Miragaia et al., 2008; Otto, 2017). In this work, we conducted a survey at the Instituto Nacional de Perinatología “INPer” in México City, over a period spanning eight years, to register changes in the antibiotic profile of staphylococci species and strains. We focus on *SE* because it was the bacteria most frequently found among the staphylococci isolates and their remarkable multi-resistance pattern. The number of antibiotic resistances remained very similar year by year with most *SE* strains multi-resistant to nine antibiotics; a minor representation of low-resistant strains (< 5 antibiotics) was found.

Moreover, the number of strains resistant to each one of the antibiotics tested remained stable through all the years. Therefore, the *SE* population within the INPer hospital was highly stable and led us to question about the genetic composition of the *SE* multi-resistant population. Specifically, we address the hypothesis that a single or few clones are the basis of the normalization of the *SE* multi-resistant strains in the children health care unit studied.

We analyzed the genomes of 17 selected *SE* INPer strains isolated mostly from neonatal patients. Our findings regarding virulence genetic determinants are concordant with those found in *SE*

isolates from hospitals and commensal strains isolated worldwide. The most prominent features of these isolates are their ability to form biofilms, the presence of PMSs, and the multidrug resistance profile displayed (Otto, 2014; Xu et al., 2018). Employing a phylogenetic strategy using the *ccrA,B,C* recombinases, we conclude that most of the INPer *SE* strains carry the SCC*mec* type IV. Eight strains of the *SE* collection analyzed here, harbor neither the *ica* operon nor the IS256, considered pathogenicity markers in *SE* (Kozitskaya et al., 2004; Murugesan et al., 2018). Likely, these are commensal *SE* strains that invaded the patients in the course of their hospital stay, even though we cannot discard they use other pathogenicity mechanisms. These strains showed resistance to multiple antibiotics, and three of them contain a composite SCC*mec* cassette formed by type IV and VIII gene elements. We propose a probable structure of the combined SCC*mec* IV and VIII cassette; however, it should be subject to further corroboration. This is not usual combination but there are reports in the literature of mosaic chromosomal staphylococcal *mec* cassettes (Heusser et al., 2007). Classification of SCC*mec* is still hard to discern due to their variability and presence of repeated elements which difficult the correct assembly of the region (International Working Group on the Classification of Staphylococcal Cassette Chromosome, 2009) The chromosomal cassettes might be hot spots of recombination of heavy metal resistance genes, insertion sequences, and antibiotic resistance genes recruited by HGT (Xue et al., 2017).

The genomic antibiotic resistance spectrum of INPer SE strains is very diverse, with some genes present in most strains and others only in few. Examples of diversification of the resistance mechanisms are the presence of membrane efflux pumps (NorAB), which may confer resistance to quinolones as well, and several modifying enzymes (AAC, APH) that inactivate aminoglycosides such as kanamycin. Indeed, the *fosB* gene, which encodes the resistance to fosfomycin was found in 9 out of 17 *SE* strains. In several of these instances, the resistance phenotype coincided with the presence of one or more genes, encoding modifying or degrading enzymes and mutant protein targets for some antibiotics (Table 2).

The genome analysis also indicates some mechanisms for antibiotic resistance, including non-synonymous substitutions in the housekeeping genes *gyrA* and *rpoB*. In the gyrase (GyrA) it was found the amino acid S84F change which confers quinolone resistance, whereas, in RpoB (β-subunit of the RNA polymerase), a double amino acid substitution D471E: I527M, and a single I527M were identified. The double mutant RpoB D471E: I527M has been recognized elsewhere as the most common cause of worldwide rifampicin resistance (Lee et al., 2018). Indeed, the presence of this RpoB variant reduces the susceptibility to vancomycin and teicoplain. In this work, the RpoB double mutant was detected in S02, S05, and S08 all belonging to ST23, while the single mutant I527M was observed in the strains S14, S17, S18, and S19, which are within the ST2. Both are the clonal lineages worldwide distributed reported by Lee et al., (2019). These INPer strains have the capacity to form biofilms and contain most of the virulence determinants (Table 1). Therefore, it is likely that they are very adapted *SE* strains present during several years in the INPer.

In the phylogenetic trees reported, pathogenic *SE* strains are intermingled with commensal *SE* strains. However, no pathogenic cluster was found (Miragaia et al., 2005; Meric et al., 2018). Recently, Meric *et al.,* suggested, that pathogenic *SE* subpopulations occur within the commensal *SE* strains, which contain genes and alleles necessary for colonization infection sites (Meric et

al., 2018). These Genome-wide-association (GWAS) studies showed the enrichment of several genes involved with methicillin resistance (*mecA*), biofilm formation *ica*), cell toxicity, and inflammation response in pathogenic *SE* isolates (Meric et al., 2018). Therefore, pathogenic and commensal *SE* strains likely concur in the same infection site, but current clinical methods of isolation prevent us from distinguishing one from the other. Eight *ica*$^-$ out of 17 analyzed, could not form biofilms even thought they were recovered from ill patients (Table 1).

There are various footprints of MGEs, including prophages, ISs, and the phage immunity CRISPR-Cas systems in the *SE* genomes that likely contribute to the adaptability by the acquisition of virulence and antibiotic resistance factors. As expected, due to its mobile nature, these elements do not follow a uniform distribution in the phylogeny, indicating frequent genetic exchange in the *SE* population. Together HGT, homologous recombination may be a factor for genetic diversification of *SE* at hospital settings. In our work, extensive genome analysis of the rates of recombination versus mutation suggests that recombination affects the whole genome and not only a particular class of genes. It has been estimated that recombination could involve 40% of the genome of *SE*, whereas in *S. aureus* recombination comprise the 24% portion (Meric et al., 2015). Although recombination rates depend strongly on the sample of strains used for the analysis, the estimated r/m values reported agrees with other recombination test performed with distinct samples of *SE*, few or whole-genome markers as well as reported r/m numbers in the literature (Miragaia et al., 2005; Meric et al., 2015). Therefore we can conclude that the *SE* population despite its whole low level of nucleotide variation (ANIm > 97%) shows cohesive clonal behavior but frequent gene exchange and recombination.

Conclusions

At local hospital settings, pathogenic and commensal *SE* strains coexist, but it is hard to discern if they are contaminants, commensal colonizers, or virulent strains (Widerstrom et al., 2016). Indeed, single-colony testing for the identification of *SE* isolates limits to know the extent of multiclonal or multi-species infections (Van Eldere et al., 2000; Harris et al., 2016). In the present study, some analyzed *SE* INPer strains came from nosocomial patients, but lack *ica* genes, a classical virulence marker. However, we cannot exclude that these putative commensal *SE* strains are pathogens in other conducts. Likely they are part of the intra-hospital non-pathogenic microbiome. These presumed *SE* commensal strains as well as the biofilm formers considered pathogenic *SE* strains are multi-resistant to antibiotics. The results present here, including an 8-year survey, suggest that the multi-resistance to antibiotics might drive adaptation and persistence of certain *SE* clones in hospital settings. HGT and recombination might play a crucial role in the origin of the pathogenic clones, moving and recombining antibiotic resistance and virulence genes in distinct genomic clonal backgrounds, including non-pathogenic strains. Therefore, the clinic and genetic factors that influence the stability and change of *SE* community overtime should be addressed in detail in future studies.

Acknowledgements

Authors recognized the Instituto Nacional de Perinatología (INPer)–for providing the *S. epidermidis* strains used in this work. Thanks to Gabriela Guerrero, Luis Lozano and José

Espíritu for computational help. We express thanks to Olga M. Pérez-Carrascal for her advice on the use of ClonalFrameML, and to Santiago Castillo for discussion on recombination analysis.

References

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at:http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antonopoulos, D.A., Assaf, R., Aziz, R.K., Brettin, T., Bun, C., Conrad, N. et al. (2017) PATRIC as a unique resource for studying antimicrobial resistance. Brief Bioinform 10.1093/bib/bbx083.

Arndt, D., Marcu, A., Liang, Y., and Wishart, D.S. (2017) PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. Brief Bioinform 10.1093/bib/bbx121.

Aswani, V.H., Tremblay, D.M., Moineau, S., and Shukla, S.K. (2014) Complete Genome Sequence of a Staphylococcus epidermidis Bacteriophage Isolated from the Anterior Nares of Humans. Genome Announc 2.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19: 455-477.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bouchami, O., de Lencastre, H., and Miragaia, M. (2016) Impact of Insertion Sequences and Recombination on the Population Structure of Staphylococcus haemolyticus. PLoS One 11: e0156653.

Burts, M.L., Williams, W.A., DeBord, K., and Missiakas, D.M. (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections. Proc Natl Acad Sci U S A 102: 1169-1174.

Byrd, A.L., Belkaid, Y., and Segre, J.A. (2018) The human skin microbiome. Nat Rev Microbiol 16: 143-155.

Capella-Gutierrez, S., and Gabaldon, T. (2013) Measuring guide-tree dependency of inferred gaps in progressive aligners. Bioinformatics 29: 1011-1017.

Chessa, D., Ganau, G., and Mazzarello, V. (2015) An overview of Staphylococcus epidermidis and Staphylococcus aureus with a focus on developing countries. J Infect Dev Ctries 9: 547-550.

Coll, P., Coque, T.M., Domínguez, M.A., Vázquez, J., Vila, J. (2005) Métodos Moleculares de tipificación epidemiológica en bacteriología. In Procedimientos en Microbiología Clínica Recomendaciones de la Sociedad Española de Enfermedade Infecciosas y Microbiología Clíni. Coll, M.A.D.n.P., Coque, M.T., Vázquez, M.A.D.n.J., and Vila, J. (eds).

Conlan, S., Mijares, L.A., Program, N.C.S., Becker, J., Blakesley, R.W., Bouffard, G.G. et al. (2012) Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome Biol 13: R64.

Contreras-Moreira, B., and Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79: 7696-7701.

Dai, Y., Wang, Y., Liu, Q., Gao, Q., Lu, H., Meng, H. et al. (2017) A Novel ESAT-6 Secretion System-Secreted Protein EsxX of Community-Associated Staphylococcus aureus Lineage ST398 Contributes to Immune Evasion and Virulence. Front Microbiol 8: 819.

539 Daniel, A., Bonnen, P.E., and Fischetti, V.A. (2007) First complete genome sequence of two
540 Staphylococcus epidermidis bacteriophages. J Bacteriol 189: 2086-2100.
541 Davis, J.J., Boisvert, S., Brettin, T., Kenyon, R.W., Mao, C., Olson, R. et al. (2016)
542 Antimicrobial Resistance Prediction in PATRIC and RAST. Sci Rep 6: 27930.
543 Didelot, X., and Wilson, D.J. (2015) ClonalFrameML: efficient inference of recombination in
544 whole bacterial genomes. PLoS Comput Biol 11: e1004041.
545 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
546 throughput. Nucleic Acids Res 32: 1792-1797.
547 Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., and Spratt, B.G. (2004) eBURST: inferring
548 patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus
549 sequence typing data. J Bacteriol 186: 1518-1530.
550 Gutierrez, D., Martinez, B., Rodriguez, A., and Garcia, P. (2012) Genomic characterization of
551 two Staphylococcus epidermidis bacteriophages with anti-biofilm potential. BMC Genomics 13:
552 228.
553 Harris, L.G., Murray, S., Pascoe, B., Bray, J., Meric, G., Mageiros, L. et al. (2016) Biofilm
554 Morphotypes and Population Structure among Staphylococcus epidermidis from Commensal and
555 Clinical Samples. PLoS One 11: e0151240.
556 Heusser, R., Ender, M., Berger-Bachi, B., and McCallum, N. (2007) Mosaic staphylococcal
557 cassette chromosome mec containing two recombinase loci and a new mec complex, B2.
558 Antimicrob Agents Chemother 51: 390-393.
559 International Working Group on the Classification of Staphylococcal Cassette Chromosome, E.
560 (2009) Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for
561 reporting novel SCCmec elements. Antimicrob Agents Chemother 53: 4961-4967.
562 Kozitskaya, S., Cho, S.H., Dietrich, K., Marre, R., Naber, K., and Ziebuhr, W. (2004) The
563 bacterial insertion sequence element IS256 occurs preferentially in nosocomial Staphylococcus
564 epidermidis isolates: association with biofilm formation and resistance to aminoglycosides.
565 Infect Immun 72: 1210-1215.
566 Kozitskaya, S., Olson, M.E., Fey, P.D., Witte, W., Ohlsen, K., and Ziebuhr, W. (2005) Clonal
567 analysis of Staphylococcus epidermidis isolates carrying or lacking biofilm-mediating genes by
568 multilocus sequence typing. J Clin Microbiol 43: 4751-4757.
569 Lee, J.Y.H., Monk, I.R., Goncalves da Silva, A., Seemann, T., Chua, K.Y.L., Kearns, A. et al.
570 (2018) Global spread of three multidrug-resistant lineages of Staphylococcus epidermidis. Nat
571 Microbiol 3: 1175-1185.
572 Letunic, I., and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new
573 developments. Nucleic Acids Res 10.1093/nar/gkz239.
574 Marchant, E.A., Boyce, G.K., Sadarangani, M., and Lavoie, P.M. (2013) Neonatal sepsis due to
575 coagulase-negative staphylococci. Clin Dev Immunol 2013: 586076.
576 Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. Nature 526: 55-61.
577 Marraffini, L.A., and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene
578 transfer in staphylococci by targeting DNA. Science 322: 1843-1845.
579 Martinez-Melendez, A., Morfin-Otero, R., Villarreal-Trevino, L., Camacho-Ortiz, A., Gonzalez-
580 Gonzalez, G., Llaca-Diaz, J. et al. (2016) Molecular epidemiology of coagulase-negative
581 bloodstream isolates: detection of Staphylococcus epidermidis ST2, ST7 and linezolid-resistant
582 ST23. Braz J Infect Dis 20: 419-428.

583 Meric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L. et al. (2015)
584 Ecological Overlap and Horizontal Gene Transfer in Staphylococcus aureus and Staphylococcus
585 epidermidis. Genome Biol Evol 7: 1313-1328.
586 Meric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B. et al. (2018) Disease-
587 associated genotypes of the commensal skin bacterium Staphylococcus epidermidis. Nat
588 Commun 9: 5034.
589 Miragaia, M., Couto, I., and de Lencastre, H. (2005) Genetic diversity among methicillin-
590 resistant Staphylococcus epidermidis (MRSE). Microb Drug Resist 11: 83-93.
591 Miragaia, M., Thomas, J.C., Couto, I., Enright, M.C., and de Lencastre, H. (2007) Inferring a
592 population structure for Staphylococcus epidermidis from multilocus sequence typing data. J
593 Bacteriol 189: 2540-2552.
594 Miragaia, M., Carrico, J.A., Thomas, J.C., Couto, I., Enright, M.C., and de Lencastre, H. (2008)
595 Comparison of molecular typing methods for characterization of Staphylococcus epidermidis:
596 proposal for clone definition. J Clin Microbiol 46: 118-129.
597 Miragaia, M., Couto, I., Pereira, S.F., Kristinsson, K.G., Westh, H., Jarlov, J.O. et al. (2002)
598 Molecular characterization of methicillin-resistant Staphylococcus epidermidis clones: evidence
599 of geographic dissemination. J Clin Microbiol 40: 430-438.
600 Miragaia, M., de Lencastre, H., Perdreau-Remington, F., Chambers, H.F., Higashi, J., Sullam,
601 P.M. et al. (2009) Genetic diversity of arginine catabolic mobile element in Staphylococcus
602 epidermidis. PLoS One 4: e7722.
603 Murugesan, S., Mani, S., Kuppusamy, I., and Krishnan, P. (2018) Role of insertion sequence
604 element is256 as a virulence marker and its association with biofilm formation among
605 methicillin-resistant Staphylococcus epidermidis from hospital and community settings in
606 Chennai, South India. Indian J Med Microbiol 36: 124-126.
607 Otto, M. (2009) Staphylococcus epidermidis--the 'accidental' pathogen. Nat Rev Microbiol 7:
608 555-567.
609 Otto, M. (2014) Phenol-soluble modulins. Int J Med Microbiol 304: 164-169.
610 Otto, M. (2017) Staphylococcus epidermidis: a major player in bacterial sepsis? Future
611 Microbiol 12: 1031-1033.
612 Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M.,
613 Mikhailova, N. et al. (2016) Uncovering Earth's virome. Nature 536: 425-430.
614 Peschel, A., and Otto, M. (2013) Phenol-soluble modulins and staphylococcal infection. Nat Rev
615 Microbiol 11: 667-673.
616 Richter, M., Rossello-Mora, R., Oliver Glockner, F., and Peplies, J. (2016) JSpeciesWS: a web
617 server for prokaryotic species circumscription based on pairwise genome comparison.
618 Bioinformatics 32: 929-931.
619 Rolo, J., Worning, P., Nielsen, J.B., Bowden, R., Bouchami, O., Damborg, P. et al. (2017)
620 Evolutionary Origin of the Staphylococcal Cassette Chromosome mec (SCCmec). Antimicrob
621 Agents Chemother 61.
622 Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006) ISfinder: the
623 reference centre for bacterial insertion sequences. Nucleic Acids Res 34: D32-36.
624 Stamatakis, A. (2015) Using RAxML to Infer Phylogenies. Curr Protoc Bioinformatics 51: 6 14
625 11-14.
626 Thomas, J.C., Vargas, M.R., Miragaia, M., Peacock, S.J., Archer, G.L., and Enright, M.C. (2007)
627 Improved multilocus sequence typing scheme for Staphylococcus epidermidis. J Clin Microbiol
628 45: 616-619.

629 Van Eldere, J., Peetermans, W.E., Struelens, M., Deplano, A., and Bobbaers, H. (2000)
630 Polyclonal Staphylococcal endocarditis caused by genetic variability. Clin Infect Dis 31: 24-30.
631 Villari, P., Sarnataro, C., and Iacuzio, L. (2000) Molecular epidemiology of Staphylococcus
632 epidermidis in a neonatal intensive care unit over a three-year period. J Clin Microbiol 38: 1740-
633 1746.
634 Vos, M., and Didelot, X. (2009) A comparison of homologous recombination rates in bacteria
635 and archaea. ISME J 3: 199-208.
636 Wang, Y., Hu, M., Liu, Q., Qin, J., Dai, Y., He, L. et al. (2016) Role of the ESAT-6 secretion
637 system in virulence of the emerging community-associated Staphylococcus aureus lineage
638 ST398. Sci Rep 6: 25163.
639 Wattam, A.R., Brettin, T., Davis, J.J., Gerdes, S., Kenyon, R., Machi, D. et al. (2018) Assembly,
640 Annotation, and Comparative Genomics in PATRIC, the All Bacterial Bioinformatics Resource
641 Center. Methods Mol Biol 1704: 79-101.
642 Widerstrom, M., McCullough, C.A., Coombs, G.W., Monsen, T., and Christiansen, K.J. (2012)
643 A multidrug-resistant Staphylococcus epidermidis clone (ST2) is an ongoing cause of hospital-
644 acquired infection in a Western Australian hospital. J Clin Microbiol 50: 2147-2151.
645 Widerstrom, M., Wistrom, J., Edebro, H., Marklund, E., Backman, M., Lindqvist, P., and
646 Monsen, T. (2016) Colonization of patients, healthcare workers, and the environment with
647 healthcare-associated Staphylococcus epidermidis genotypes in an intensive care unit: a
648 prospective observational cohort study. BMC Infect Dis 16: 743.
649 Xu, Z., Misra, R., Jamrozy, D., Paterson, G.K., Cutler, R.R., Holmes, M.A. et al. (2018) Whole
650 Genome Sequence and Comparative Genomics Analysis of Multi-drug Resistant Environmental
651 Staphylococcus epidermidis ST59. G3 (Bethesda) 8: 2225-2230.
652 Xue, H., Wu, Z., Qiao, D., Tong, C., and Zhao, X. (2017) Global acquisition of genetic material
653 from different bacteria into the staphylococcal cassette chromosome elements of a
654 Staphylococcus epidermidis isolate. Int J Antimicrob Agents 50: 581-587.
655 Zerbino, D.R. (2010) Using the Velvet de novo assembler for short-read sequencing
656 technologies. Curr Protoc Bioinformatics Chapter 11: Unit 11 15.

657

Figure 1

Profiles of *SE* antibiotic multi-resistance in a eight years period

A-H, number of antibiotic resistances of 595 *SE*isolates by year. I-O, number of strains resistant to each one of fourteen antibiotics by year. The *SE*strains selected for genome analysis are indicated over the bars.

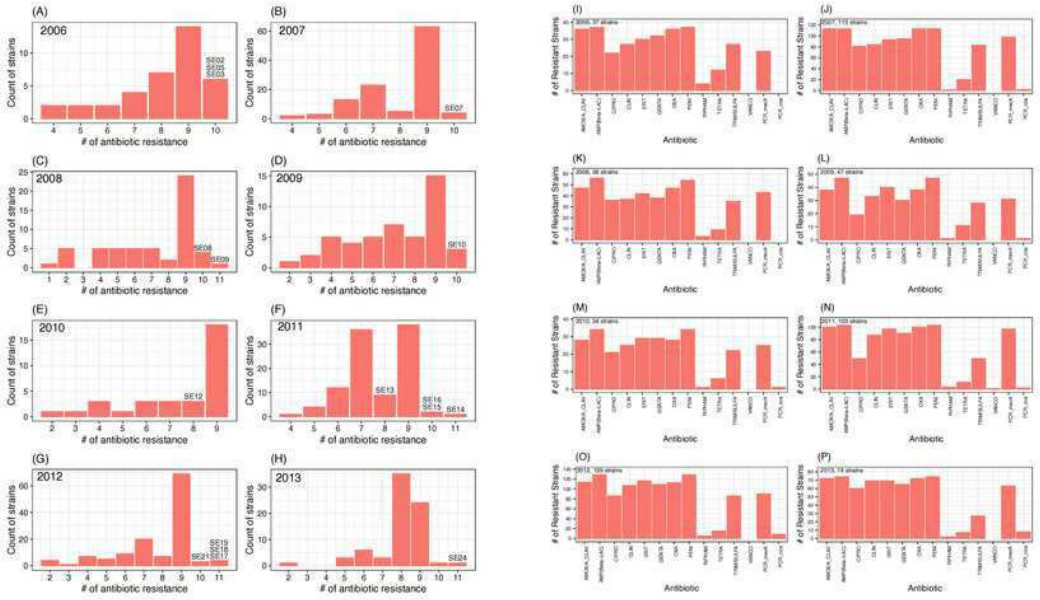

Figure 2

Phylogenetic relationships of *SE* INPer strains and *SE* selected from GenBank

The ML tree consists of four main clades defined by the longest branches (A to D). They are indicated with color dots: red, clade A; blue, clade B; green, clade C; yellow, clade D. The tree was constructed with 1575 common core proteins using RaxML program as described in methods. The results of bootstrap performed with 1000 replicates are indicated in the branches. Acronyms specify the isolation site of the *SE* strains: MEX, México; USA, United States of America; FIN, Finland; GER, Germany; DEN, Denmark; IRE, Ireland; CHI, China; AUS, Australia. *SE* INPer strains are denoted in red. Asterisks indicate the strains that lack of the *ica*operon. All the other *SE*INPer strains present the *ica;* reference strains were not included in *ica* searches.

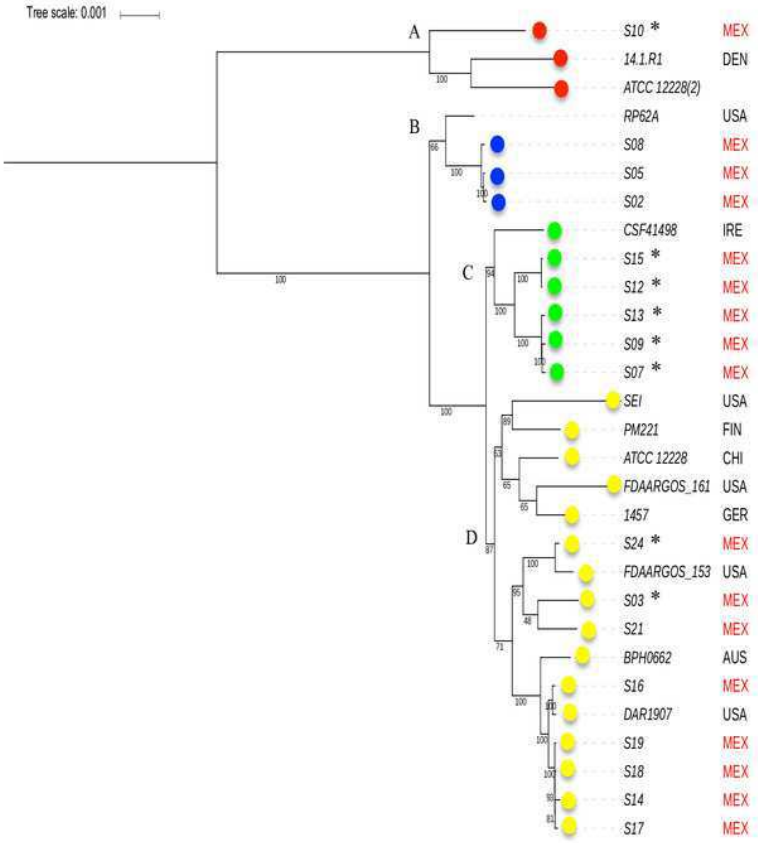

Figure 3

Distribution of accessory genes in the *SE* genomes.

The heat-map profile was performed with the ggplot2 function in R using the Bray-Curtiss dissimilarity matrix. The Bray-Curtiss dendrogram is indicated at the left. The heat-map at the middle indicates gene presence (blue color); empty cells represent the absence of genes. *SE s*trains are shown at the left with the same colors of the clades in the core proteins phylogeny (Fig. 1): red, clade A; blue, clade B; green, clade C; yellow, clade D. In the last column. the presence/absence of genomic regions similar to known prophages is indicated (Table S4).

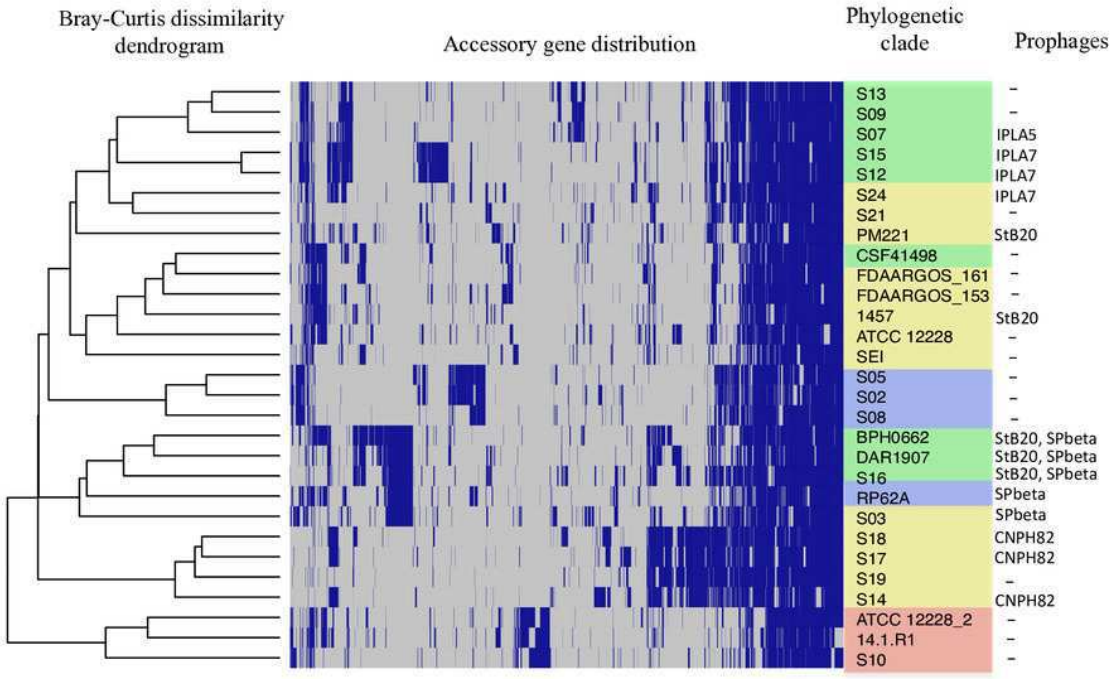

Figure 4

Rates of recombination/mutation for *SE* INPer strains compared with sets of *SE* from GenBank.

1. Sixteen strains out 17 *SE* of the INPer collection; 2. Seventeen *SE* strains of the INPer collection. 3- 22, r/m rates for the genes encoding proteins classified in COGs in the 17 *SE* INPer: 3. COG C (energy production). 4. COG D (cell division). 5. COG E (amino acid transport and metabolism); 6. COG F (nucleotide transport and metabolism). 7. COG G (carbohydrate transport and metabolism). 8. COG H (coenzyme transport and metabolism). 9. COG I (lipid transport and metabolism). 10. COG J (translation, ribosomal structure and biogenesis). 11. COG K (transcription). 12. COG L (replication, recombination and repair). 13. COG M (cell wall, membrane, envelope biogenesis). 14. COG O (post-translational modification, protein turnover and chaperones). 15. COG P (inorganic ion transport and metabolism). 16. COG Q (secondary metabolites biosynthesis, transport and catabolism). 17. COG R (general function predicted). 18. COG S (function unknown). 19. COG T (signal transduction mechanisms). 20. COG U (intracellular trafficking, secretion and vesicular transport). 21. COG V (defense mechanism). 22. Unassigned COGs. 23. Antibiotic resistance genes predicted in PATRIC server for the *SE* INPer strains. 24. Virulence genes predicted in PATRIC server for the *SE* INPer strains. 25. Reference set of 12 complete genomes of *SE* strains used through this work. 26-33, subsets of draft *SE* strains from GenBank: 26. Ctr1 (n =36). 27. Ctr2 (n =26). 28. Ctr3 (n =22). 29. Ctr4 (n =35). 30. Ctr-5 (n =36). 31. Ctr6 (n =26). 32. Ctr7 (n =36). 33. Ctr8 (n =36). Descriptions of the *SE* strains included in the control sets and their GenBank accession numbers are in Table S3.

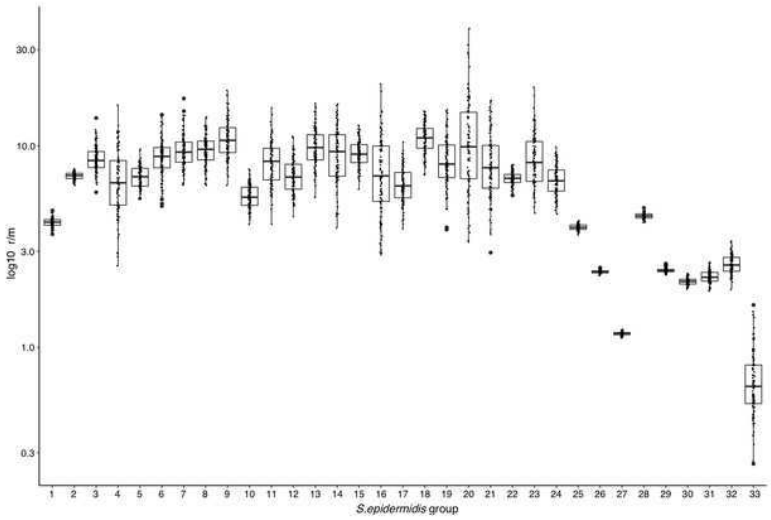

Figure 5

Genome-wide recombination between *SE* strains.

ClonalFrameML program detected several events of recombination along 1,354,455 concatenated genomic regions of 29 *SE* genomes (17 INPer genomes and 12 *SE*GenBank genomes). RaxML nucleotide tree is shown at the left of the scheme. Color dots indicate the corresponding clades in the protein phylogeny (Figure 1). Blue bars indicate recombination events along the concatenated genome segments. White bars indicate non-homoplasic nucleotide substitutions; yellow to red bars are probable homoplasic nucleotide substitutions (Didelot and Wilson, 2015). Red dots line indicates ancestral events of recombination.

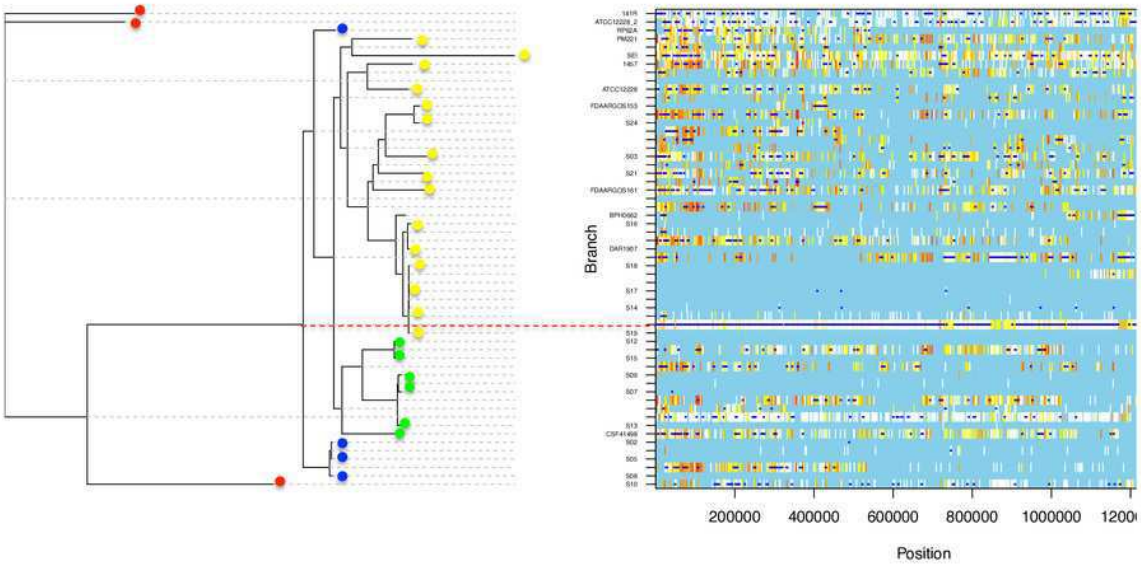

Table 1(on next page)

General features of the genomes of *SE* INPer strains.

[a] S05 lack of *icaC*; [d] N prefix indicates strains isolated from neonates; A, isolates from adults; CSF, cerebrospinal fluid; STA, soft tissue aspirate. [c]U-ST, unassigned ST.[d] ST profile numbers correspond to alleles of ArcC, AroE, Gtr, MutS, TpiA, and YqiL, according to Thomas et al. (2007). [e]Parentheses indicate the number of complete IS256. NA, indicate the presence of a *ccrC* gene but not assigned to any SCC*mec* type. SCC*mec* classification was performed by phylogenetic clustering of the recombinases *ccrA, ccrB, and ccrC* with the recombinases present in the SCC*mec* types.

									SCCmec cassette type				
Strain [a]	**Origin** [b]	**ST Type** [c]	**ST profile** [d]	***icaADBRC***	**CRISPR-Cas**	**Prophages**	**IS256** [e]	***mecA***	***ccrA***		***ccrB***		***ccrC***
S02	N-Catheter	23	7,1,2,1,3,3,1	+	1	–	1 (1)	+	IV		IV		IV
S03	N-Catheter	89	1,1,2,1,2,1,1	–	0	+	1 (1)	+	IV		IV		
S05	N-Catheter	23	7,1,2,1,3,3,1	+	1	–	1 (1)	+	IV		IV		IV
S07	A-STA	59	2,1,1,1,2,1,1	–	0	+	–	+	IV	VIII	IV	VIII	
S08	N-Blood	23	7,1,2,1,3,3,1	+	1	–	1 (1)	+	IV		IV		
S09	A-Blood culture	59	2,1,1,1,2,1,1	–	1	–	–	+	IV	VIII	IV	VIII	
S10	A-Blood	U-ST	5,48,5,5,8,5,1	–	0	–	–	–					
S12	N-Blood	5	1,1,1,2,2,1,1	–	3	+	1 (1)	+	IV		IV		IV
S13	N-Blood culture	81	2,17,1,1,2,1,1	–	1	–	1(0)	+	IV	VIII	IV	VIII	NA
S14	N-Blood culture	2	7,1,2,2,4,1,1	+	0	+	2 (1)	+	IV		IV		
S15	N-Catheter	5	1,1,1,2,2,1,1	–	0	+	1 (1)	+	IV		IV		NA
S16	N-Blood culture	2	7,1,2,2,4,1,1	+	0	+	1 (1)	–					
S17	N-Catheter	2	7,1,2,2,4,1,1	+	0		2 (1)	+	IV		IV		
S18	N-Blood	2	7,1,2,2,4,1,1	+	0	+	2 (1)	+	IV		IV		
S19	N-Catheter	2	7,1,2,2,4,1,1	+	0	–	2 (1)	+	IV		IV		
S21	N-Catheter	35	2,1,2,2,4,1,1	+	1	–	–	+					NA
S24	N-CSF	5	1,1,1,2,2,1,1	–	3	+	–	+	IV		IV		

Table 2(on next page)

Antibiotic resistance phenotypes and the genotype profile in *SE* INPer strains

Resistant strains are indicated by yellow cells. Susceptible are in red cells. Empty cells indicate strains not tested for the respective antibiotic. *Parentheses in the column indicates the number of strains that present the most probable gene (s) involved in antibiotic resistance. In *gyrA* and *rpoB* the corresponding amino acid substitution in the protein is indicated.

		S. epidermidis strains																Genotype *
Antibiotic	**Class**	**S2**	**S03**	**S05**	**S07**	**S08**	**S09**	**S10**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**	**S18**	**S19**	**S21**	**S24**	Genes probably involved
Ampicillin	Penicillin																		*blaZ* (17), *mecA* (15)
Penicillin	Penicillin																		*blaZ* (17), *mecA* (15)
Amoxicillin	Penicillin																		*blaZ* (17), *mecA* (15)
Oxacillin	Penicillin																		*blaZ* (17), *mecA* (15)
Imipenem	Carbapenem																		*blaZ* (17), *mecA* (15)
Cefalotin	Cefalosphorin																		*blaZ* (17), *mecA* (15)
Cefazolin	Cefalosphorin																		*blaZ* (17), *mecA* (15)
Cefoxitin	Cefalosphorin																		*blaZ* (17), *mecA* (15)
Ceftriaxone	Cefalosphorin																		*blaZ* (17), *mecA* (15)
Ciprofloxacin	Quinolone																		*gyrA* S84F (10)
Moxifloxacin	Quinolone																		*gyrA* S84F (10)
Levofloxacin	Quinolone																		*gyrA* S84F (10)
Clarithromycin	Macrolide																		*mphB* (3), *ermC* (11)
Erythromycin	Macrolide																		*mphB* (3), *ermC* (11)
Clindamycin	Lyncosamide																		*linA* (1), *ermC* (11)
Tetracycline	Tetracicline																		*tetL* (4)
Gentamycin	Aminoglycoside																		*aac/aph* (15), *aph3* (2), *ant9* (3), *aadD* (11)
Trimethoprim/Sulfamethoxazole	Sulfonamina																		*sul3* (17)
Chloramphenicol	Phenicol																		*catB* (6)
Rifampicin	Miscellaneus																		*rpoB* D471E, I527M (3); I527M (4)
Vancomycin	Glycopeptide																		*vanRI* (17)