Published in final edited form as:

Infect Genet Evol. 2014 March ; 22: 257–264. doi:10.1016/j.meegid.2013.06.020.

Differing lifestyles of *Staphylococcus epidermidis* as revealed through Bayesian clustering of multilocus sequence types

Jonathan C. Thomas[1], **Liangfen Zhang**[1], and **D. Ashley Robinson**[1,*]

[1]Department of Microbiology, University of Mississippi Medical Center, Jackson, MS, USA

Abstract

Staphylococcus epidermidis is part of the normal bacterial flora of human skin and a leading cause of infections associated with indwelling medical devices. Previous phylogenetic analyses of subgenomic data have been unable to distinguish between *S. epidermidis* strains with nosocomial or commensal lifestyles, despite the identification of specific phenotypes and accessory genes that may contribute to such lifestyles. To attempt to better define the population structure of this species, the international *S. epidermidis* multilocus sequence typing database was analyzed with the Bayesian clustering programs STRUCTURE and BAPS. A total of six genetic clusters (GCs) were identified. A local population of *S. epidermidis* from clinical specimens was classified according to these six GCs, and further characterized for antibiotic susceptibilities, biofilm, and various genetic markers. GC5 was abundant and significantly enriched for isolates that were resistant to four classes of antibiotics, high biofilm production, and positive for the virulence markers *icaA*, IS*256*, and *sesD*/*bhp*, indicating its potential clinical relevance. In contrast, GC2 was rare and contained the only isolates positive for the putative commensal marker, *fdh*. GC1 and GC6 were abundant but not significantly associated with any of the examined characteristics, except for *sesF/aap* and GC6. GC3 was rare and identified as a potential genetic sink that received, but did not donate, core genetic material from other GCs. In conclusion, population genetics analyses were essential for identifying clusters of strains that may differ in their adaptation to nosocomial or commensal lifestyles. These results provide a new, population genetics framework for studying *S. epidermidis*.

Keywords

Bayesian clustering; population assignment; population structure; *Staphylococcus epidermidis*; multilocus sequence typing

1. INTRODUCTION

Staphylococcus epidermidis is a ubiquitous resident of the human skin. However, over the past 30 years, this species has emerged as an important opportunistic pathogen that causes infections associated with indwelling medical devices such as intravenous catheters and prosthetic heart valves (Fang et al., 1993; Richet et al., 1990; Zandri et al., 2012). *S. epidermidis*, along with other coagulase-negative staphylococci, is now a leading cause of

*Corresponding author: Mailing address: Department of Microbiology, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216. Phone: (601) 984-1702. Fax: (601) 984-1708. darobinson@umc.edu.

nosocomial bloodstream infections (Luzzaro et al., 2011). In neonates and immunocompromised patients, *S. epidermidis* is more prevalent as a cause of infection than its more pathogenic relative, *S. aureus* (Villari et al., 2000).

S. epidermidis virulence factors have been identified through comparisons of nosocomial isolates with isolates colonizing the skin of healthy individuals in the community and healthy individuals with varying degrees of hospital contact (Kozitskaya et al., 2005; Rohde et al., 2004; Rolo et al., 2012). Some *S. epidermidis* isolates appear to be high biofilm producers and, hence, may better persist on catheter tips and other indwelling medical devices (Gill et al., 2005; Mekni et al., 2012). Many *S. epidermidis* virulence factors are involved in biofilm formation, such as the production of the polysaccharide intercellular adhesin encoded by the *ica* locus (Heilmann et al., 1996), and the accumulation-associated protein encoded by *sesF*/*aap* (Rohde et al., 2005). In addition, IS*256* has been associated with isolates from nosocomial infections (Gu et al., 2005; Kozitskaya et al., 2004; Yao et al., 2005), and *fdh*, which encodes a formate dehydrogenase, has been associated with commensal isolates (Conlan et al., 2012).

While *S. epidermidis* may not carry a wide array of virulence factors, the species has become increasingly antibiotic-resistant; up to 50% of isolates were multidrug-resistant in some studies (Mendes et al., 2012). Oxacillin resistance has been reported in up to 70% of nosocomial isolates (Flamm et al., 2013; Mendes et al., 2012). Indeed, Rolo et al. (2012) have noted that *S. epidermidis* isolates from the hospital environment more frequently carry an SCC*mec* genetic element, which harbours the *mecA* gene that is responsible for broad β-lactam resistance.

The population structure of *S. epidermidis* has not been fully revealed. eBURST analysis of multilocus sequence typing (MLST) data places the majority of *S. epidermidis* sequence types (STs) into one clonal complex (Miragaia et al., 2007; Wisplinghoff et al., 2003; Wong et al., 2010), which confounds attempts to identify lineages that might be associated with more or less virulent lifestyles. Lineages that are enriched for isolates with virulence or commensal markers have not been identified by earlier studies that have used less portable, but potentially more discriminatory, strain typing tools (Begović et al., 2013; Kozitskaya et al., 2005; Rohde et al., 2004). Some data support the hypothesis that relatively high rates of recombination occur between strains of *S. epidermidis* (Kozitskaya et al., 2005; Miragaia et al., 2007), which may impede the differentiation of lineages, but this hypothesis requires further testing with data that is more powerful than MLST data (Zhang et al., 2012).

Difficulties in defining population structure for *S. epidermidis* may arise from an insufficient number of informative polymorphisms and the confounding effects of recombination, but there is also the broader challenge of defining a bacterial “population” (Robinson et al., 2011). Multiple definitions of a population have been proposed for non-bacterial species (Waples and Gaggiotti, 2006). For bacterial species with relatively low rates of recombination, populations may be defined by their clonal lineages (Rannala et al., 2000). However, for species such as *S. epidermidis* that may be more recombinant, Bayesian model-based clustering may be used to define populations. The Bayesian clustering programs STRUCTURE and BAPS have been applied to study the population structures of a diverse array of bacterial species such as *Burkholderia pseudomallei*, *Chlamydia trachomatis*, *Enterococcus faecium*, *Escherichia coli*, *Helicobacter pylori*, *Salmonella enterica* and *Streptococcus pneumoniae* (Dale et al., 2011; Didelot et al., 2011; Falush et al., 2003b; Gordon et al., 2008; Hanage et al., 2009; Joseph et al., 2012; Willems et al., 2012). In both programs, the strain composition of *K* clusters, where *K* is defined *a priori,* is determined essentially by either minimizing the within-cluster genetic diversity or maximizing the between-cluster genetic diversity, while allowing for admixture to occur

between clusters. Here, both STRUCTURE and BAPS were used to elucidate the population structure of *S. epidermidis*, at global and local scales. In addition, the distribution of antibiotic susceptibilities, biofilm, and a variety of genetic markers was examined, which leads to a new understanding of *S. epidermidis* population structure and how it may be related to the bacteria's lifestyle.

2. MATERIALS AND METHODS

2.1 Bacterial Isolates

The isolate collection used in this study was described previously by Smyth et al. (2011). Briefly, for a 6-month period between January and June 2007, all staphylococcal isolates from clinical specimens were collected on a weekly basis from the Microbiology Laboratory of Westchester Medical Center (WMC) in Valhalla, NY. The 129 *S. epidermidis* isolates used here were previously typed by MLST (Zhang et al., 2012), according to published methods for *S. epidermidis* (Thomas et al., 2007). Bacteria were grown overnight on tryptic soy agar plates at 37°C. Isolates were stored long-term at −80°C in tryptic soy broth (TSB) with 15% glycerol (v/v). Bacterial genomic DNA was extracted with a DNeasy kit (Qiagen), as per the manufacturer's instructions. Characteristics of all study isolates are listed in Supplemental Table 1.

2.2 Antibiotic Susceptibilities

Susceptibilities to clindamycin, erythromycin, gentamicin, oxacillin and trimethoprim-sulfamethoxazole (TMP-SMX) were determined at the time of isolation by MicroScan Gram-positive MIC susceptibility panel and a MicroScan WalkAway system (Dade Behring, Inc.), or by broth microdilution. All susceptibilities were checked by disk diffusion tests for this study, with the exception of oxacillin that had been checked previously (Smyth et al., 2011). Susceptibilities were intepreted according to the Clinical and Laboratory Standards Institute (CLSI, 2007). To detect inducible clindamycin resistance, "D-tests" were performed by placing an erythromycin disk 15 mm from the clindamycin disk. A positive D-test was indicated by a flattening of the zone of inhibition bordering the erythromycin disk (CLSI, 2007). Isolates positive for inducible clindamycin resistance were considered to be resistant for the purposes of analysis.

2.3 Biofilm Assay

The degree of biofilm formation of each *S. epidermidis* isolate was determined using the method of Christensen et al. (1985) with the modifications of O'Neill et al. (2007). In brief, isolates were randomly assigned to Nunclon™ ΔSurface, flat-bottomed, 96-well polystyrene plates and grown in brain heart infusion (BHI) medium at 37°C. Optical density was measured at 600 nm after 24 hours growth. Plates were rinsed with sterile distilled water four times. Plates were subsequently dried at 60°C for 1 hour, stained with a 0.4% crystal violet solution, and optical density at 492 nm was determined with a Bio-Rad xMark™ microplate spectrophotometer. *S. epidermidis* strain RP62a was a positive control and sterile BHI medium was a negative control on each plate. Three technical replicates were performed for each isolate per plate, and three biological replicates for each isolate were performed on separate plates to provide the average optical density for each isolate (A_{OD}). Due to the negative correlation between biofilm production ($A_{OD}492$) and growth ($A_{OD}600$), as described in Results section 3.6, biofilm values for each isolate were first normalized to their respective growth values, before being normalized to the positive control for each plate. Biofilm values were then averaged across the three biological replicates. High biofilm production was defined as an $A_{OD}492$ greater than the median $A_{OD}492$ for all isolates.

2.4 Detection of Various Genetic Markers

Isolates had been screened previously by PCR for the presence of the arginine catabolic mobile element (ACME) (Zhang et al., 2012). Here, all isolates were screened by PCR for the presence of the virulence markers *icaA*, IS*256* and *mecA*, as well as the surface protein-encoding genes *sesA*-*I*, and the putative commensal marker *fdh*. The genes *sesD* and *sesF* are also known as *bhp* and *aap*, respectively (Bowden et al., 2005; Gill et al., 2005). PCR was repeated for all isolates where a gene was variably present within an ST, and where results for three or fewer STs differed from the remainder of their genetic cluster. Primers and PCR conditions are listed in Supplemental Table 2.

2.5 Bayesian and Non-Bayesian Clustering

Allelic and sequence data for all 437 sequence types (STs) from the international MLST database as of 29th June 2012 were assigned to clonal complexes using eBURST (Feil et al., 2004) and goeBURST (Francisco et al., 2009) and to genetic clusters (GCs) using the Bayesian clustering programs STRUCTURE v2.3.3 (Falush et al., 2003a) and BAPS v5.4 (Corander et al., 2008). For the two BURST analyses, which were based on allelic data, all STs in a clonal complex differed by no more than one allele from at least one other ST in the clonal complex.

Sequence data were analyzed with STRUCTURE using the admixture model with correlated allele frequencies. Each run consisted of a Monte Carlo Markov Chain (MCMC) of 200,000 iterations, with the first 100,000 iterations discarded as burn-in and the subsequent 100,000 iterations saved. Each analysis was independently repeated five times. The number of clusters present in the sample (*K*) was varied between 1–11 and the resulting estimated logs of the posterior probabilities of the data (Ln P(D)) for each *K* were averaged. Following Evanno et al. (2005), Δ*K* was calculated for *K*>2, and used to determine the optimal number of clusters. An individual ST was assigned to a cluster provided that the proportion of its ancestry (Q) derived from that cluster was >0.7. For comparison, the no-admixture and linkage models were also performed, and all models were performed separately with the correlated and independent allele frequency assumptions.

For analyses with BAPS, sequences for each locus were trimmed, orientated, and concatenated to maintain the +1 reading frame. The program's module for clustering with linked loci was utilized with a codon linkage model. The maximum number of clusters was varied between 11–20, and the analyses were independently repeated five times. Population admixture analysis was performed with the mixture clustering module, using 100 iterations to estimate admixture coefficients. The admixture significance threshold was set to α=0.05 and "gene flow pathways" representing less than 0.01 (1%) donated genetic material were pruned.

The relationships between global *S. epidermidis* BAPS-defined GCs were inferred from a neighbor-joining analysis of D_A-distances calculated from single nucleotide polymorphism (SNP) frequencies, using the POPTREE2 program (Takezaki et al., 2010). The relationships between global *S. epidermidis* STs were inferred from a neighbor-joining analysis of *p*-distances calculated from the concatenated MLST sequences, using the program MEGA v5.05 (Tamura et al., 2011). Support for both tree topologies was assessed with 1000 bootstrap replicates. F_{ST} between genetic clusters was calculated using DnaSP v5.10 (Librado and Rozas, 2009).

2.6 Statistical Analyses

The adjusted Rand index (ARI; Hubert and Arabie, 1985) was used to measure the concordance between the different classifications of STs into clonal complexes and GCs.

ARI calculations using cluster assignments from STRUCTURE were performed without those STs that could not be assigned to a cluster. Confidence intervals around the ARI point estimate were calculated using a jackknife method (Smyth et al., 2011; Severiano et al., 2011). Odds ratios were used to compare the proportions of significantly admixed STs among each of the GCs detected in the global *S. epidermidis* population. The strength of the associations between GCs and various markers was tested by Fisher's exact tests or chi-squared tests as appropriate, using InStat v3.1 (GraphPad Software). Spearman's correlation between biofilm production ($A_{OD}492$) values and growth ($A_{OD}600$) values was calculated using Prism v6.0 (GraphPad Software).

3. RESULTS

3.1 Genetic Clusters of a Global Population of *S. epidermidis*

The global population of *S. epidermidis*, as sampled by the international MLST database, consisted of 437 sequence types (STs) as of 29th June 2012. eBURST placed the 437 STs into 16 clonal complexes of more than 2 STs each, and 96 singletons, although only four of the clonal complexes contained more than five STs and 271 (62%) STs belonged to a single clonal complex (Supplemental Fig. 1). goeBURST identified exactly the same groups as eBURST (Supplemental Fig. 2). BAPS analysis identified six genetic clusters (GCs), ranging in size from 20 (4.6%) to 137 (31%) STs (Fig. 1A). The admixture model of STRUCTURE detected seven GCs based on Evanno et al.'s (2005) statistic of ΔK (Supplemental Fig. 3A, 3B). Since no STs derived mainly (Q>0.7) from two of these seven GCs, the STs belonged to five GCs as defined by STRUCTURE, one of which corresponded to a combined GC2 and GC4, as defined by BAPS (Fig. 1B).

3.2 Concordance Between Clustering Methods

The adjusted Rand index (ARI) was used to measure the concordance of cluster assignments from different STRUCTURE models. Treatment of allele frequencies as correlated or independent generally had little impact on the concordance of different models (Supplemental Table 3; ARI≥0.930). The admixture and linkage models both allow for population admixture, and these models were relatively concordant with each other (ARI≥0.869). The no-admixture model with independent allele frequencies was the least concordant model overall (ARI≥0.763).

The ARI of the cluster assignments from STRUCTURE and BAPS was 0.699 (95% CI 0.639, 0.760); however, other than the collapse of BAPS GC2 and GC4 into a single cluster by STRUCTURE, the only other differences were the 70 STs of indeterminate cluster by STRUCTURE (Q<0.7), 35 of which corresponded to STs that were identified as significantly admixed by BAPS. After removing these 70 STs, the ARI was 0.910 (95% CI 0.876, 0.944), which was a significantly higher concordance. By comparison, the clusters identified by STRUCTURE and BAPS differed significantly from the clonal complexes identified by eBURST. This finding was robust to whether the STs defined as singletons by eBURST were excluded, treated as a single group, or treated as separate distinct groups (Supplemental Table 4; ARI≤0.251). The Bayesian-defined GCs distinguish between STs that may be inappropriately linked in eBURST-defined clonal complexes; the coloring in Supplemental Fig. 1 and 2 indicates how BAPS-defined GCs break up clonal complexes. Due to the considerable agreement between the results from STRUCTURE and BAPS, the remainder of this study utilized the more rapidly computed global BAPS assignments, which consisted of six GCs.

3.3 Relationships Between Genetic Clusters

A close relationship between GCs 2 and 4, and between GCs 1 and 6, was supported by strong bootstrap support in the population tree (Fig. 2A). Bootstrap support for the positions of GC3 and GC5 in the population tree was ≤66%. Similar relationships between GCs were observed from a tree based on the concatenated MLST sequences of the 437 STs (Fig. 2B); however, some mixing of STs between GCs was also apparent on this tree.

3.4 Admixture Between Genetic Clusters

BAPS was used to assess the degree of admixture in the global *S. epidermidis* population and to plot the gene flow between GCs. In total, 58 STs distributed across all GCs in the global population were determined to be significantly admixed ($P<0.05$) by BAPS. The majority of STs belonging to GC1 possessed sequence from only that cluster, and GC1 contained significantly fewer admixed STs than the overall population (Table 1). GC2, GC4 and GC6 did not significantly differ from the overall population in terms of the proportion of admixed STs ($P>0.05$). In contrast, GC3 and GC5 each contained significantly more admixed STs than the overall population (Table 1); 21% of GC3's nucleotides, and 6% of GC5's nucleotides, were admixed (Fig. 3). The gene flow graph also showed that GC3 received sequence from all other GCs, whereas GC5 received sequence from only GCs 2 and 6 (Fig. 3). The greater amount of admixture in GC3 compared to other GCs is also consistent with its weaker bootstrap support in the population tree (Fig. 2A) and the apparent polyphyly of its STs (Fig. 2B).

3.5 Genetic Clusters of a Local Population of *S. epidermidis*

A local population of *S. epidermidis*, isolated from clinical specimens and previously typed by MLST, consisted of 129 isolates and 29 STs. When this local population was classified according to the eBURST-defined clonal complexes from the global population, all but 5 isolates belonged to a single clonal complex (data not shown). In contrast, when this local population was classified according to the BAPS-defined GCs from the global population, five of the six GCs were present. The majority of the isolates belonged to GC1 (n=31), GC5 (n=44) and GC6 (n=49), whereas GC2 (n=2) and GC3 (n=3) were relatively rare and GC4 was absent in this local population.

Interestingly, separate analyses of the local population of *S. epidermidis* with STRUCTURE and BAPS identified only two clusters (data not shown), demonstrating that strain sampling strongly influences the number of clusters identified. Concordance between the global and local cluster assignments was relatively low for both STRUCTURE and BAPS, with ARI of 0.380 (95% CI 0.200, 0.569) and 0.479 (95% CI 0.358, 0.608), respectively. This was due to both programs collapsing GC2 and GC5 into a single cluster, and collapsing GC1 and GC6 into a single cluster. The three GC3 isolates were of indeterminate cluster by STRUCTURE, while in BAPS they formed a third cluster that would have been discounted based on the small sample size (n=3; Latch et al., 2006).

3.6 Statistical Associations Between Genetic Clusters and Virulence and Commensal Markers

Using data from 128 isolates (one isolate from GC1 repeatedly failed to grow during biofilm assay) that were normalized to the positive control for each plate, average biofilm production ($A_{OD}492$) values were negatively correlated with growth ($A_{OD}600$) values (Spearman's $r=-0.45$, $P<0.0001$; Fig. 4A). An additional normalization of each isolate's biofilm value to its respective growth value was therefore warranted, since biofilm production could be affected by how well the isolate grew. Isolates with an $A_{OD}492$ greater

than the median value of 0.053, were judged to be high biofilm producers (Fig. 4B). This classification of biofilm was used in subsequent statistical analyses.

Significant, non-random associations were found by chi-squared tests between GC and resistance to clindamycin, gentamicin, oxacillin and TMP-SMX, high biofilm production, and presence of ACME, *icaA*, IS*256*, *sesD*/*bhp* and *sesF*/*aap* ($P<0.05$; Table 2). By Fisher's exact test, GC5 was significantly enriched for isolates resistant to clindamycin, gentamicin, oxacillin and TMP-SMX, high biofilm production, and the presence of *icaA*, IS*256* and *sesD*/*bhp* ($P<0.02$). Conversely, GC5 was significantly underrepresented for isolates positive for ACME ($P=0.013$) and *sesF*/*aap* ($P=0.0002$). GC6 was significantly enriched for isolates positive for *sesF*/*aap* ($P=0.0006$). Several of the *S. epidermidis* surface protein-encoding genes were either very common (e.g. *sesA*, *sesB*, *sesC*, *sesE* and *sesH*) or very rare (e.g. *sesG* and *sesI*) and therefore not amenable to these statistical analyses (Table 2). Despite the significant, non-random association between GC and resistance to oxacillin, no association was found between GC and presence of *mecA*, possibly due to the disagreement of genotypic and phenotypic data for ten isolates (Supplemental Table 1). However, this discrepancy has been described previously by several studies of *S. epidermidis* (Martinaeu et al., 2000; McDonald et al., 1995; Tenover et al., 1999).

4. DISCUSSION

A Bayesian clustering approach has been used previously to determine population structure in various bacterial species (Dale et al., 2011; Didelot et al., 2011; Falush et al., 2003b; Gordon et al., 2008; Hanage et al., 2009; Joseph et al., 2012; Willems et al., 2012). Before the development of this approach, bacterial "populations" were sometimes defined *a priori* based on characteristics that could be easily observed, such as host species or geography, and the genetic differentiation between populations was then tested against a null hypothesis of no differentiation (Selander et al., 1986; Whittam et al., 1989). However, Bayesian clustering makes it possible to first define bacterial populations based on their genetic differentiation, and then determine the characteristics associated with each population (Robinson et al., 2011).

Here, genetic clusters (GCs) were identified in the global *S. epidermidis* population by two Bayesian clustering programs; five GCs were identified by STRUCTURE, and six GCs were identified by BAPS, although two of the clusters identified by BAPS, GC2 and GC4, were collapsed into a single cluster by STRUCTURE. Separate analyses of the local *S. epidermidis* population, identified only two clusters even though five of the six GCs were present, highlighting the role of strain sampling in cluster identification. A reference list of ST-GC mappings for all STs in the MLST database as of 29th June 2012 is provided in Supplemental Table 5. Even though these GCs were identified from only seven loci, a relatively high degree of genetic differentiation was detected between these GCs (average $F_{ST}=0.61$). Simulation studies have found that both STRUCTURE and BAPS are able to accurately define the number of clusters in a species where F_{ST} is as low as 0.03 (Latch et al., 2006), which indicates there is sufficient structure in the global MLST data to be confident in the performance of these programs in this study.

GC5 appears to be suited to a more nosocomial lifestyle; it is significantly associated with resistance to several antibiotics, biofilm, and virulence markers such as *icaA*, IS*256*, and *sesD*/*bhp*, which may contribute to survival in a nosocomial environment. GC5 was the second most commonly isolated cluster (34%) from the clinical specimens of our local *S. epidermidis* population. Our finding of GC5's association with multiple virulence markers is consistent with several recent studies that have reported ST2 and ST23, both of which belong to GC5, as being frequently isolated from clinical specimens (Iorio et al., 2012; Li et

al., 2009; Mendes et al., 2012; Widerström et al., 2012). GC5 was also underrepresented by isolates that are positive for ACME and *sesF*/*aap*. Studies of ACME have tended to focus on its potential virulence role in *S. aureus*, although this particular role is still unclear (Diep et al., 2008; Joshi et al., 2011; Montgomery et al., 2009). In *S. epidermidis*, ACME has been reported to be highly prevalent among carriage isolates (Barbier et al., 2011) and associated with low levels of antibiotic resistance and pathogenicity (Granslo et al., 2010).

GCs 2 and 4 appear to be suited to a more commensal lifestyle. The group of commensal isolates identified in the genomic analysis of Conlan et al. (2012) corresponds to GC4 in the global *S. epidermidis* population; these isolates were obtained from skin swabs from healthy volunteers. Conlan et al. (2012) identified the gene *fdh* from these GC4 isolates as a potential commensal-associated marker. Our results suggest that both GC2 and GC4 may have a more commensal lifestyle; GC2 and GC4 were collapsed into a single cluster by STRUCTURE, and they were closely related in the population tree, while *fdh* was present only in the two GC2 isolates from our local population. Although our local *S. epidermidis* isolates lack specific clinical information other than their isolation from clinical specimens, the rarity of GC2 and absence of GC4 in our local population supports the notion that their niche is not the clinical environment.

GCs 1 and 6 appear to have a more generalist lifestyle. Alternatively, these two GCs may be adapted to a nosocomial lifestyle, but possess virulence factors that have not been recognized as such in *S. epidermidis*. In our local *S. epidermidis* population, isolates from these two GCs were not associated with the examined markers, except for *sesF*/*aap* and GC6, yet these isolates were prevalent among the clinical specimens. A recent study by Rolo et al. (2012) described 24 STs that were present in both hospital and community environments in their study. Of those 24 STs, 20 (83%) belonged to GC1 and GC6, indicating that these GCs are capable of surviving in both hospital and community environments. Thus, our conclusion that GCs 1 and 6 have a more generalist lifestyle is also consistent with the results of Rolo et al. (2012).

Admixture analysis of the global *S. epidermidis* population identified asymmetric gene flow between GCs. In particular, GC3 was the most admixed cluster and was the recipient of gene flow from the five other clusters, but it was not the source of gene flow to any of the other clusters. Therefore, GC3 may be described as a genetic sink; that is, a population that imports genetic material at a higher rate than it exports genetic material (Pulliam, 1988). Bacterial populations that behave in such a manner may acquire a diversity of mobile genetic elements and other genetic material external to the species (Caro-Quintero et al., 2011). MLST data for only 20 isolates of GC3 is present in the international database, including the data from the three isolates of our local population, so GC3 may be relatively rare or under-sampled.

The MLST database represents the most complete catalog of *S. epidermidis* genetic variants at present, but we do not presume that it includes examples of all *S. epidermidis* lineages. Since the improved MLST scheme for *S. epidermidis* was released in 2007 (Thomas et al., 2007), no isolates from Oceania have been deposited in the database, and only 53 isolates from Africa or Asia have been deposited. Consequently, there is potentially a very broad range of *S. epidermidis* diversity that has not yet been characterized. Sparse sampling of some lineages can cause those lineages to appear “recombinant” in Bayesian clustering analyses (Denamur et al., 2010; Gordon et al., 2008; Wirth et al., 2006). Given the rarity of GC3 and its recombinant characteristics, genotyping of isolates from geographic regions where MLST of *S. epidermidis* has been sparse will be necessary to confirm that GC3 is a single admixed cluster rather than multiple distinct clusters. The sequences from MLST loci represent a very small portion of the *S. epidermidis* genome, but they detected adequate

differentiation between GCs. However, complete genome sequences of isolates from each GC should be able to validate the clusters identified here, provided that recombination is accounted for, and they may provide further biological information about *S. epidermidis* lifestyles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grant GM080602 from the National Institutes of Health (to D.A.R.).

References

Barbier F, Lebeaux D, Hernandez D, Delannoy AS, Caro V, François P, Schrenzel J, Ruppé E, Gaillard K, Wolff M, Brisse S, Andremont A, Ruimy R. High prevalence of the arginine catabolic element in carriage isolates of methicillin-resistant *Staphylococcus epidermidis*. J Antimicrob Chemother. 2011; 66:29–36. [PubMed: 21062794]

Begović J, Jovčić B, Papić-Obradović M, Veljović K, Lukić J, Kojić M, Topisirović L. Genotypic diversity and virulent factors of *Staphylococcus epidermidis* isolated from human breast milk. Microbiol Res. 2013; 168:77–83. [PubMed: 23098640]

Bowden MG, Chen W, Singvall J, Xu Y, Peacock SJ, Valtulina V, Speziale P, Höök M. Identification and preliminary characterization of cell-wall-anchored proteins of *Staphylococcus epidermidis*. Microbiology. 2005; 151:1453–1464. [PubMed: 15870455]

Caro-Quintero A, Deng J, Auchtung J, Brettar I, Höfle MG, Klappenbach J, Konstantinidis KT. Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. ISME J. 2011; 5:131–140. [PubMed: 20596068]

Christensen GD, Simpson WA, Younger JJ, Baddour LM, Barrett FF, Melton DM, Beachey EH. Adherence of coagulase-negative staphylococci to plastic tissue culture plates: a quantitative model for the adherence of staphylococci to medical devices. J Clin Microbiol. 1985; 22:996–1006. [PubMed: 3905855]

Clinical and Laboratory Standards Institute – CLSI. Performance standards for antimicrobial susceptibility testing. Wayne, PA: 2007. p. M100-S17

Conlan S, Mijares LA, Becker J, Blakesley RR, Bouffard GG, Brooks S, Coleman HL, Gupta J, Gurson N, Park M, Schmidt B, Thomas PJ, Young A, Otto M, Kong HH, Murray PR, Segre JA. NISC Comparative Sequencing Program. *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome Biol. 2012; 13:R64. [PubMed: 22830599]

Corander J, Marttinen P, Siren J, Tang J. Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. BMC Bioinformatics. 2008; 9:539. [PubMed: 19087322]

Dale J, Price EP, Hornstra H, Busch JD, Mayo M, Godoy D, Wuthiekanun V, Baker A, Foster JT, Wagner DM, Tuanyok A, Warner J, Spratt BG, Peacock SJ, Currie BJ, Keim P, Pearson T. Epidemiological tracking and population assignment of the non-clonal bacterium, *Burkholderia pseudomallei*. PLoS Negl Trop Dis. 2011; 5:e1381. [PubMed: 22180792]

Denamur, E.; Picard, B.; Tenaillon, O. Population genetics of *Escherichia coli*. In: Robinson, DA.; Falush, D.; Feil, EJ., editors. Bacterial population genetics in infectious disease. Wiley-Blackwell; Hoboken, NJ: 2010. p. 269-286.

Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P. Recombination and population structure in *Salmonella enterica*. PLoS Genet. 2011; 7:e1002191. [PubMed: 21829375]

Diep BA, Stone GG, Basuino L, Graber CJ, Miller A, des Etages SA, Jones A, Palazzolo-Ballance AM, Perdreau-Remington F, Sensabaugh GF, DeLeo FR, Chambers HF. The arginine catabolic mobile element and staphylococcal chromosomal cassette mec linkage: convergence of virulence

and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. J Infect Dis. 2008; 197:1523–1530. [PubMed: 18700257]

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005; 14:2611–2620. [PubMed: 15969739]

Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data; linked loci and correlated allele frequencies. Genetics. 2003a; 164:1567–1587. [PubMed: 12930761]

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003b; 299:1582–1585. [PubMed: 12624269]

Fang G, Keys TF, Gentry LO, Harris AA, Rivera N, Getz K, Fuchs PC, Gustafson M, Wong ES, Goetz A, Wagener MM, Yu VL. Prosthetic valve endocarditis resulting from nosocomial bacteremia. A prospective, multicenter study. Ann Intern Med. 1993; 199:560–567. [PubMed: 8363166]

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004; 186:1518–1530. [PubMed: 14973027]

Flamm RK, Mendes RE, Ross JE, Sader HS, Jones RN. An international activity and spectrum analysis of linezolid: ZAAPS Program results for 2011. Diag Microbiol Infect Dis. 2013; 76:206–213.

Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. 2009; 10:152. [PubMed: 19450271]

Gill SR, Fouts DE, Archer GL, Mongodin EF, DeBoy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol. 2005; 187:2426–2438. [PubMed: 15774886]

Gordon DM, Clermont O, Tolley H, Denamur E. Assigning *Escherichia coli* strains to phylogenetic groups: multilocus sequence typing versus the PCR triplex method. Environ Microbiol. 2008; 10:2484–2496. [PubMed: 18518895]

Granslo HN, Klingenberg C, Fredheim EG, Rønnestad A, Mollnes TE, Flaegstad T. Arginine catabolic mobile element is associated with low antibiotic resistance and low pathogenicity in *Staphylococcus epidermidis* from neonates. Pediatr Res. 2010; 68:237–241. [PubMed: 20703143]

Gu J, Li H, Vuong C, Otto M, Wen Y, Gao Q. Bacterial insertion sequence IS*256* as a potential molecular marker to discriminate invasive strains from commensal strains of *Staphylococcus epidermidis*. J Hosp Infect. 2005; 61:342–348. [PubMed: 16242209]

Hanage WP, Fraser C, Tang J, Connor TR, Corander J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science. 2009; 324:1454–1457. [PubMed: 19520963]

Heilmann C, Schweitzer O, Gerke C, Vanittanakom N, Mack D, Götz F. Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. Mol Microbiol. 1996; 20:1083–1091. [PubMed: 8809760]

Hubert L, Arabie P. Comparing partitions. J Classif. 1985; 2:193–218.

Iorio NL, Caboclo RF, Azevedo MB, Barcellos AG, Neves FP, Domingues RM, dos Santos KR. Characteristics related to antimicrobial resistance and biofilm formation of widespread methicillin-resistant *Staphylococcus epidermidis* ST2 and ST23 lineages in Rio de Janeiro hospitals, Brazil. Diagn Microbiol Infect Dis. 2012; 72:32–40. [PubMed: 22100013]

Joseph SJ, Didelot X, Rothschild J, de Vries HJC, Morré SA, Read TD, Dean D. Population genomics of *Chlamydia trachomatis*: Insights on drift, selection, recombination and population structure. Mol Biol Evol. 2012; 29:3933–3946. [PubMed: 22891032]

Joshi GS, Spontak JS, Klapper DG, Richardson AR. Arginine catabolic mobile element encoded *speG* abrogates the unique hypersensitivity of *Staphylococcus aureus* to exogenous poly-amines. Mol Microbiol. 2011; 82:19–20.

Kondo Y, Ito T, Ma XX, Watanabe S, Kreiswirth BN, Etienne J, Hiramatsu K. Combination of multiplex PCRs for staphylococcal cassette chromosome *mec* type assignment: rapid identification system for *mec*, *ccr*, and major differences in junkyard regions. 2007; 51:264–274.

Kozitskaya S, Cho SH, Dietrich K, Marre R, Naber K, Ziebuhr W. The bacterial insertion sequence element IS256 occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides. Infect Immun. 2004; 72:1210–1215. [PubMed: 14742578]

Kozitskaya S, Olson ME, Fey PD, Witte W, Ohlsen K, Ziebuhr W. Clonal analysis of *Staphylococcus epidermidis* isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. J Clin Microbiol. 2005; 43:4751–4757. [PubMed: 16145137]

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE Jr. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conserv Genet. 2006; 7:295–302.

Li M, Wang X, Gao Q, Lu Y. Molecular characterization of *Staphylococcus epidermidis* strains isolated from a teaching hospital in Shanghai, China. J Med Microbiol. 2009; 58:456–461. [PubMed: 19273641]

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]

Loader C. Local likelihood density estimation. Ann Stat. 1996; 24:1602–1618.

Luzzaro F, Ortisi G, Larosa M, Drago M, Brigante G, Gesu G. Prevalence and epidemiology of microbial pathogens causing bloodstream infections: results of the OASIS multicenter study. Diagn Microbiol Infect Dis. 2011; 69:363–369. [PubMed: 21396530]

Martineau F, Picard FJ, Lansac N, Ménard C, Roy PH, Ouellette M, Bergeron MG. Correlation between the resistance genotype determined by multiplex PCR assays and the antibiotic susceptibility patterns of *Staphylococcus aureus* and *Staphylococcus epidermidis*. Antimicrob Agents Chemother. 2000; 44:231–238. [PubMed: 10639342]

McDonald CL, Maher WE, Fass RJ. Revised interpretation of oxacillin MICs for *Staphylococcus epidermidis* based on *mecA* detection. Antimicrob Agents Chemother. 1995; 39:982–984. [PubMed: 7786008]

Mekni MA, Bouchami O, Achour W, Hassen AB. Strong biofilm production but not adhesion virulence factors can discriminate between invasive and commensal *Staphylococcus epidermidis* strains. APMIS. 2012; 120:605–611. [PubMed: 22779682]

Mendes RE, Deshpande LM, Costello AJ, Farrell DJ. Molecular epidemiology of *Staphylococcus epidermidis* clinical isolates from U.S. hospitals. Antimicrob Agents Chemother. 2012; 56:4656–4661. [PubMed: 22687512]

Miragaia M, Thomas JC, Couto I, Enright MC, de Lencastre H. Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. J Bacteriol. 2007; 189:2540–2552. [PubMed: 17220222]

Montgomery CP, Boyle-Vavra S, Daum RS. The arginine catabolic mobile element is not associated with enhanced virulence in experimental invasive disease caused by the community-associated methicillin-resistant *Staphylococcus aureus* USA300 genetic background. Infect Immun. 2009; 77:2650–2656. [PubMed: 19380473]

O'Neill E, Pozzi C, Houston P, Smyth D, Humphreys H, Robinson DA, O'Gara JP. Association between methicillin susceptibility and biofilm regulation in *Staphylococcus aureus* isolates from device-related infections. J Clin Microbiol. 2007; 45:1379–1388. [PubMed: 17329452]

Pulliam HR. Sources, sinks, and population regulation. Am Nat. 1988; 132:652–661.

Rannala B, Qiu WG, Dykhuizen DE. Methods for estimating gene frequencies and detecting selection in bacterial populations. Genetics. 2000; 155:499–508. [PubMed: 10835376]

Richet H, Hubert B, Nitemberg G, Andremont A, Buu-Hoi A, Ourbak P, Galicier C, Veron M, Boisivon A, Bouvier AM, Ricome JC, Wolff MA, Pean Y, Berardi-Grassias L, Bourdain JL, Hautefort B, Laaban JP, Tillant D. Prospective multicenter study of vascular-catheter-related

complications and risk factors for positive central-catheter cultures in intensive care unit patients. J Clin Microbiol. 1990; 28:2520–2525. [PubMed: 2254429]

Robinson, DA.; Thomas, JC.; Hanage, WP. Population structure of pathogenic bacteria. In: Tibayrenc, M., editor. Genetics and evolution of infectious diseases. London: Elsevier; 2011. p. 43-57.

Rohde H, Kalitzky M, Kröger N, Scherpe S, Horstkotte MA, Knobloch JK, Zander AR, Mack D. Detection of virulence-associated genes not useful for discriminating between invasive and commensal *Staphylococcus epidermidis* strains from a bone marrow transplant unit. J Clin Microbiol. 2004; 42:5614–5619. [PubMed: 15583290]

Rohde H, Burdelski C, Bartscht K, Hussain M, Buck F, Horstkotte MA, Knobloch JK, Heilmann C, Hermann M, Mack D. Induction of *Staphylococcus* biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. Mol Microbiol. 2005; 55:1883–1895. [PubMed: 15752207]

Rolo J, de Lencastre H, Miragaia M. Strategies of adaptation of *Staphylococcus epidermidis* to hospital and community: amplification and diversification of SCC*mec*. J Antimicrob Chemother. 2012; 67:1333–1341. [PubMed: 22422509]

Selander RK, Korhonen TK, Vaisanen-Rhen V, Williams PH, Pattison PE, Caugant DA. Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect Immun. 1986; 52:213–222. [PubMed: 2870026]

Severiano A, Carriço JA, Robinson DA, Ramirez M, Pinto FR. Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PLoS One. 2011; 6:e19539. [PubMed: 21611165]

Smyth DS, Wong A, Robinson DA. Cross-species spread of SCC*mec* IV subtypes in staphylococci. Infect Genet Evol. 2011; 11:446–453. [PubMed: 21172458]

Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol. 2010; 27:747–752. [PubMed: 20022889]

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28:2731–2739. [PubMed: 21546353]

Tenover FC, Jones RN, Swenson JM, Zimmer B, McAllister S, Jorgensen JH. Methods for improved detection of oxacillin resistance in coagulase-negative Staphylococci: Results of a multicenter study. J Clin Microbiol. 1999; 37:4051–4058. [PubMed: 10565931]

Thomas JC, Vargas MR, Miragaia M, Peacock SJ, Archer GL, Enright MC. Improved Multilocus sequence typing scheme for *Staphylococcus epidermidis*. J Clin Microbiol. 2007; 45:616–619. [PubMed: 17151213]

Villari P, Sarnataro C, Iacuzio L. Molecular epidemiology of *Staphylococcus epidermidis* in a neonatal intensive care unit over a three-year period. J Clin Microbiol. 2000; 38:1740–1746. [PubMed: 10790091]

Waples RS, Gaggiotti O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 2006; 15:1419–1439. [PubMed: 16629801]

Whittam TS, Wolfe ML, Wilson RA. Genetic relationships between *Escherichia coli* isolates causing urinary tract infections in humans and animals. Epidemiol Infect. 1989; 102:37–46. [PubMed: 2645153]

Widerström M, McCullough CA, Coombs GW, Monsen T, Christiansen KJ. A multidrug-resistant *Staphylococcus epidermidis* clone (ST2) is an ongoing cause of hospital-acquired infection in a Western Australian hospital. J Clin Microbiol. 2012; 50:2147–2151. [PubMed: 22442320]

Willems RJ, Top J, van Schaik W, Leavis H, Bonten M, Sirén J, Hanage WP, Corander J. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. mBio. 2012; 3:e00151–12. [PubMed: 22807567]

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol. 2006; 60:1136–1151. [PubMed: 16689791]

Wisplinghoff H, Rosato AE, Enright MC, Noto M, Craig W, Archer GL. Related clones containing SCC*mec* type IV predominate among clinically significant *Staphylococcus epidermidis* isolates. Antimicrob Agents Chemother. 2003; 47:3574–3579. [PubMed: 14576120]

Wong A, Reddy SP, Smyth DS, Aguero-Rosenfeld ME, Sakoulas G, Robinson DA. Polyphyletic emergence of linezolid-resistant staphylococci in the United States. Antimicrob Agents Chemother. 2010; 54:742–748. [PubMed: 19933808]

Yao Y, Sturdevant DE, Villaruz A, Xu L, Gao Q, Otto M. Factors characterizing *Staphylococcus epidermidis* invasiveness determined by comparative genomics. Infect Immun. 2005; 73:1856–1860. [PubMed: 15731088]

Zandri G, Pasquaroli S, Vignaroli C, Talevi S, Manso E, Donelli G, Biavasco F. Detection of viable but non-culturable staphylococci in biofilms from central venous catheters negative on standard microbiological assays. Clin Microbiol Infect. 2012; 18:E259–261. [PubMed: 22578149]

Zhang L, Thomas JC, Didelot X, Robinson DA. Molecular signatures identify a candidate target of balancing selection in an *arcD*-like gene of *Staphylococcus epidermidis*. J Mol Evol. 2012; 75:43–54. [PubMed: 23053194]

HIGHLIGHTS

- *Staphylococcus epidermidis* is a common, opportunistic pathogen
- Previous subgenomic phylogenetic analyses have not distinguished strain lifestyles
- Bayesian clustering reveals genetic clusters with differing lifestyles

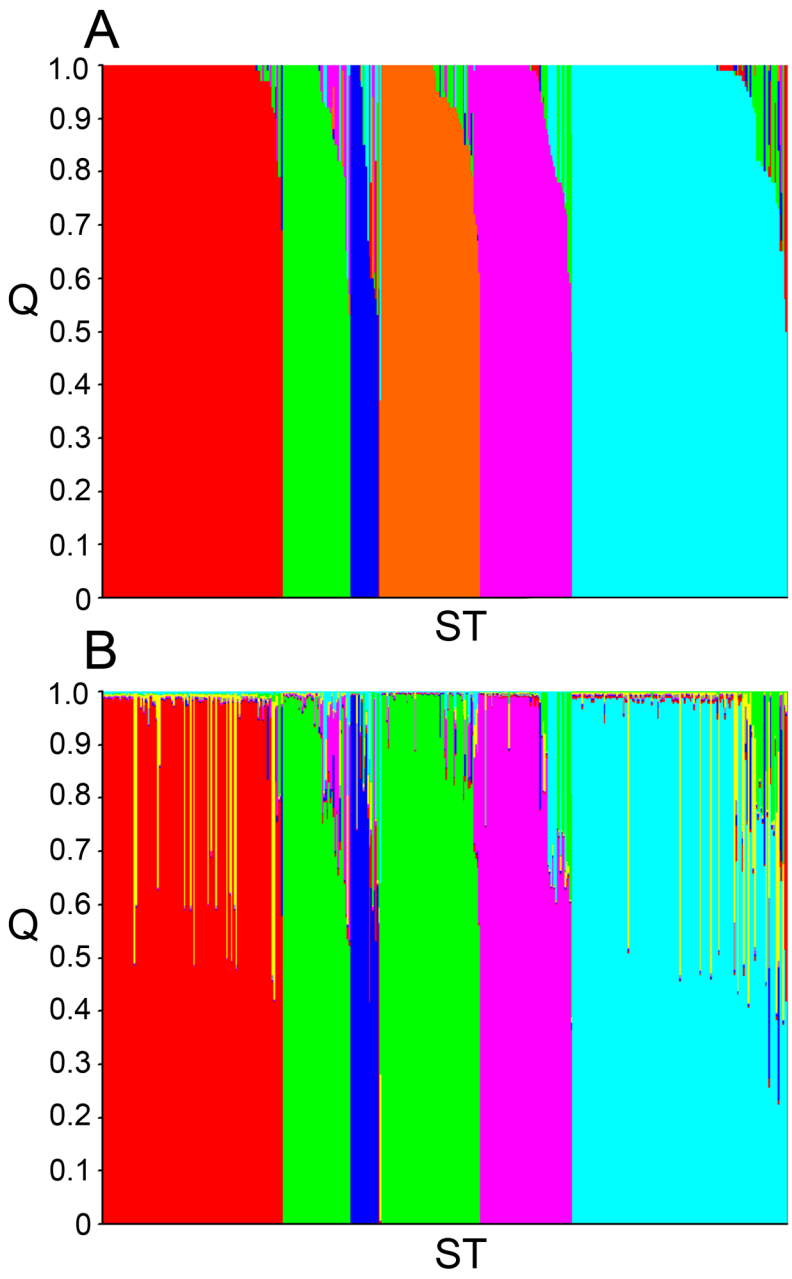

Fig. 1.

Proportion of each *S. epidermidis* sequence types' (STs) ancestry (Q) that originates from the genetic clusters (GCs) as defined by A: BAPS and B: STRUCTURE. The order of STs is the same in both panels. Red represents GC1, green represents GC2, blue represents GC3, orange represents GC4, pink represents GC5 and turquoise represents GC6.

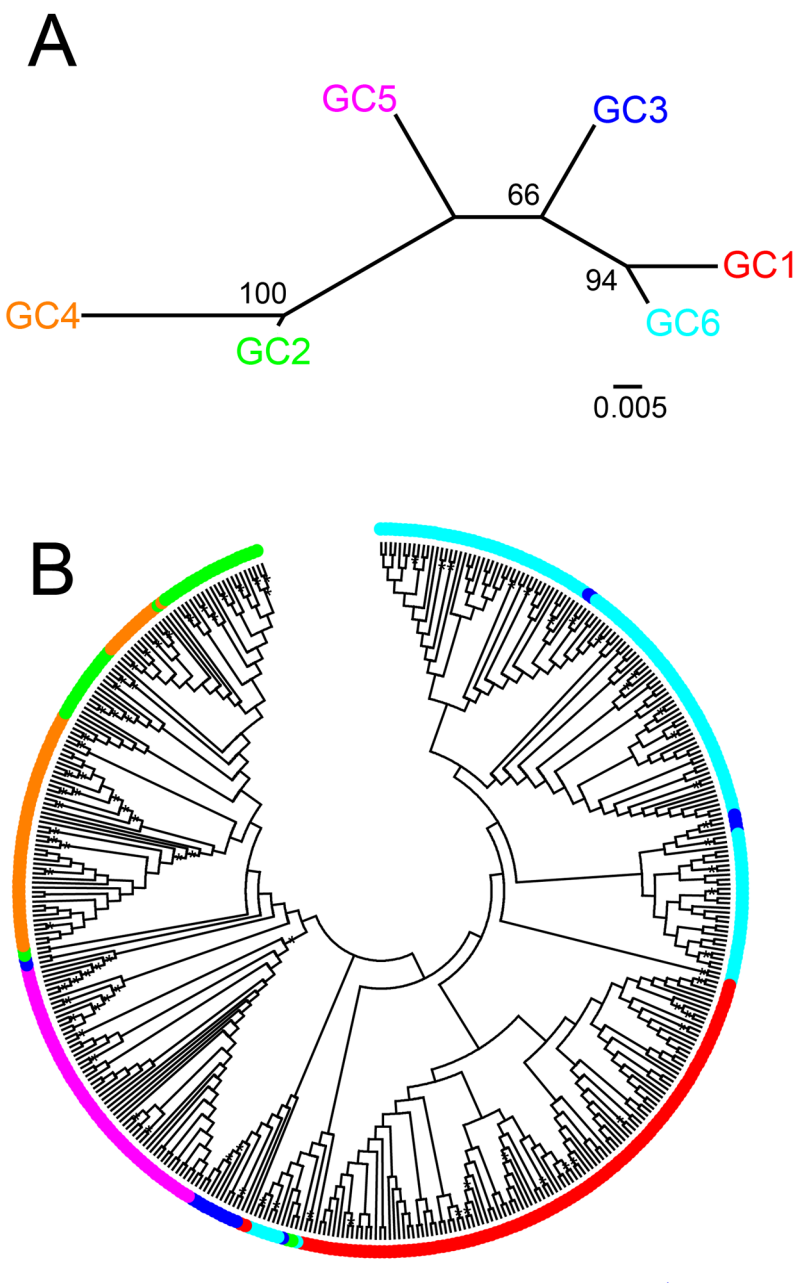

Fig. 2.

A: Population tree from neighbor-joining analysis of D_A-distances calculated from the SNP frequencies of the genetic clusters (GCs). Bootstrap support >50% is shown. B: Sequence type (ST) tree from neighbor-joining analysis of *p*-distances calculated from the concatenated MLST sequences of 437 STs. Bootstrap support >50% is shown as asterisks. Colors are as Fig. 1.

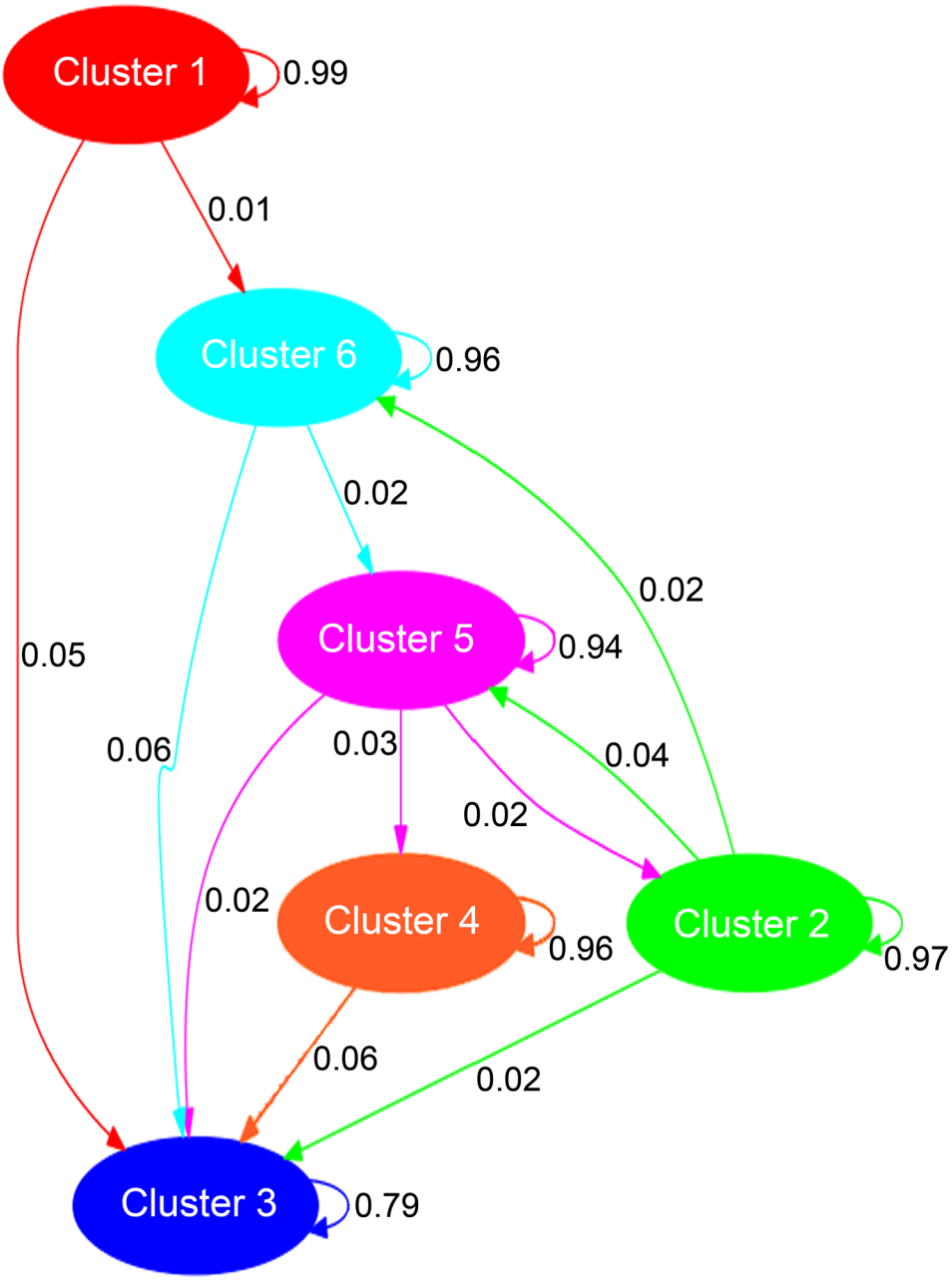

Fig. 3.
Gene flow graph depicting admixture between *S. epidermidis* genetic clusters (GCs). Colors are as Fig. 1.

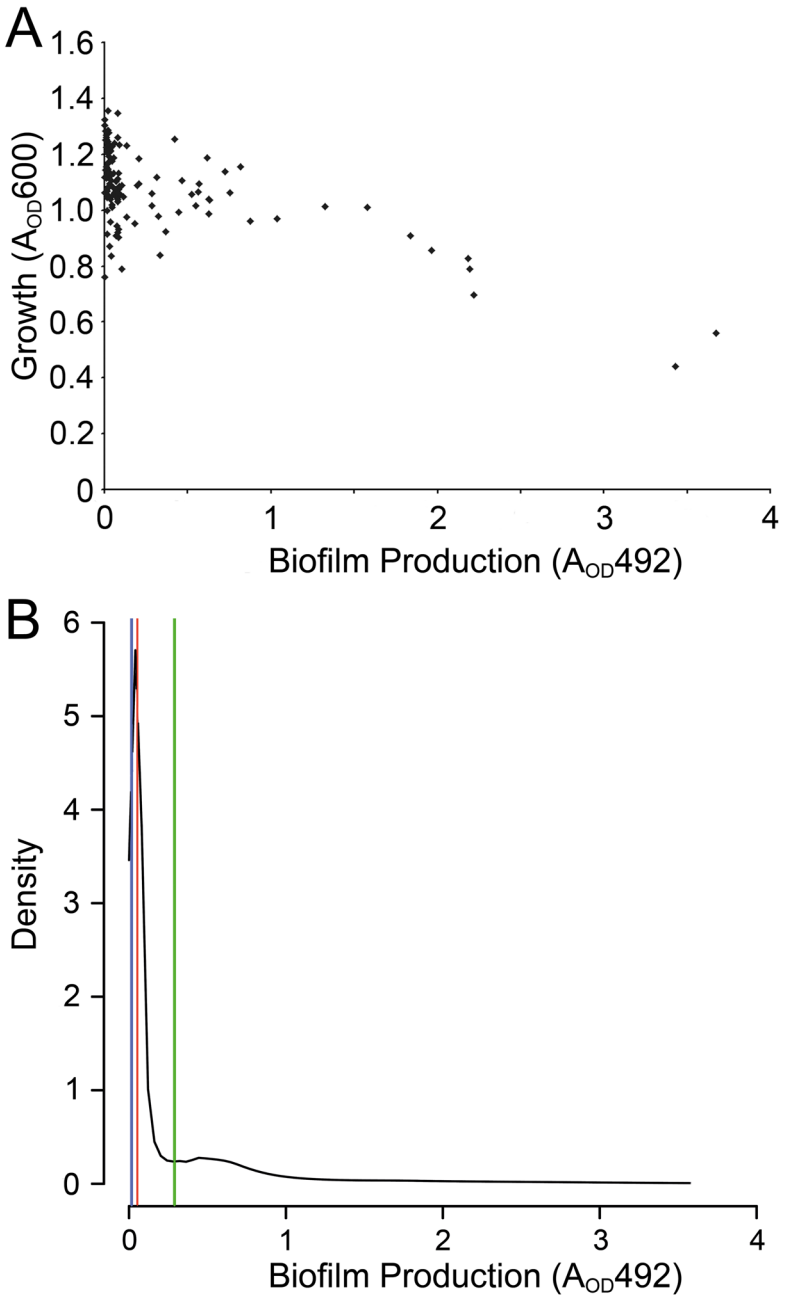

Fig. 4.

A: Relationship between biofilm production ($A_{OD}492$) values and growth ($A_{OD}600$) values; B: Distribution of biofilm production ($A_{OD}492$) values, smoothed with the default kernel of the Locfit v1.5–6 module (Loader, 1996) of the R v2.13.1 software package. The blue, red and green lines indicate the mode, median and mean values, respectively.

Table 1

BAPS admixture analysis of global *S. epidermidis* genetic clusters.

GC	No. of significantly admixed STs (total no. of STs)	Odds ratio (95% confidence interval)
1	4 (115)	**0.179 (0.063, 0.506)**
2	3 (43)	0.462 (0.138, 1.546)
3	7 (20)	**3.864 (1.473, 10.137)**
4	10 (63)	1.281 (0.611, 2.687)
5	14 (59)	**2.362 (1.200, 4.648)**
6	20 (137)	1.179 (0.658, 2.113)

Genetic cluster (GC); Sequence type (ST)

Underlined indicates that the GC is significantly enriched or underrepresented for admixed STs.

Table 2

Associations of genetic clusters with various characteristics of *S. epidermidis* from a local population.

	No. of isolates (%) positive for characteristic[a]					
Characteristic	**GC1 (n=31)**	**GC2 (n=2)**	**GC3 (n=3)**	**GC5 (n=44)**	**GC6 (n=49)**	**χ^2 (*P*-value)[b]**
High biofilm production	15 (50)	0 (0)	2 (67)	31 (70)	16 (33)	13.29 (*P*=0.0013)
icaA	20 (65)	0 (0)	2 (67)	43 (98)	15 (31)	44.80 (*P*<0.0001)
Surface protein-encoding genes						
sesA	31 (100)	2 (100)	3 (100)	44 (100)	49 (100)	N/A
sesB	31 (100)	2 (100)	3 (100)	43 (98)	49 (100)	N/A
sesC	31 (100)	2 (100)	3 (100)	42 (95)	45 (92)	N/A
sesD (*bhp*)	0 (0)	0 (0)	0 (0)	17 (39)	4 (8)	23.73 (*P*<0.0001)
sesE	31 (100)	1 (50)	3 (100)	44 (100)	49 (100)	N/A
sesF (*aap*)	25 (81)	0 (0)	2 (67)	26 (59)	46 (94)	16.61 (*P*=0.0002)
sesG	4 (13)	0 (0)	0 (0)	0 (0)	2 (4)	N/A
sesH	31 (100)	1 (50)	3 (100)	43 (98)	49 (100)	N/A
sesI	5 (16)	0 (0)	0 (0)	4 (9)	0 (0)	N/A
Antibiotic resistance						
Clindamycin	15 (48)	2 (100)	0 (0)	36 (82)	21 (43)	11.66 (*P*=0.003)
Erythromycin	26 (84)	2 (100)	0 (0)	36 (82)	32 (65)	4.91 (*P*=0.0858)
Gentamicin	7 (23)	1 (50)	0 (0)	25 (57)	14 (29)	11.66 (*P*=0.0029)
Oxacillin	25 (81)	1 (50)	0 (0)	43 (98)	42 (86)	6.35 (*P*=0.0492)
Trimethoprim-sulfamethoxazole	10 (48)	1 (50)	0 (0)	34 (77)	15 (31)	24.13 (*P*<0.0001)
mecA	26 (84)	2 (100)	0 (0)	41 (93)	46 (94)	2.71 (*P*=0.2583)
Other						
ACME	16 (52)	0 (0)	1 (33)	12 (27)	25 (51)	6.67 (*P*=0.0356)
fdh	0 (0)	2 (100)	0 (0)	0 (0)	0 (0)	N/A
IS*256*	11 (35)	2 (100)	0 (0)	38 (86)	8 (16)	48.03 (*P*<0.0001)

[a] Genetic cluster (GC); Sequence type (ST); only 30 GC1 isolates available for biofilm assay.

[b] Chi-squared *P*-values are based on the three most abundant GCs (GC1, GC5, GC6). N/A indicates the Chi-squared test was not applicable.