



HAL
open science

High-Resolution Typing of *Staphylococcus epidermidis* Based on Core Genome Multilocus Sequence Typing To Investigate the Hospital Spread of Multidrug-Resistant Clones

Anne Jamet, Julien Guglielmini, Bryan Brancotte, Mathieu Coureuil, Daniel Euphrasie, Julie Meyer, Johanna Roux, Jean-Philippe Barnier, Emmanuelle Bille, Agnès Ferroni, et al.

► To cite this version:

Anne Jamet, Julien Guglielmini, Bryan Brancotte, Mathieu Coureuil, Daniel Euphrasie, et al.. High-Resolution Typing of *Staphylococcus epidermidis* Based on Core Genome Multilocus Sequence Typing To Investigate the Hospital Spread of Multidrug-Resistant Clones. *Journal of Clinical Microbiology*, 2021, 59 (3), 10.1128/JCM.02454-20 . hal-03218144

HAL Id: hal-03218144

<https://hal.science/hal-03218144>

Submitted on 5 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **High resolution typing of *Staphylococcus epidermidis* based on cgMLST to investigate**
2 **the hospital spread of multidrug resistant clones**

3
4 **Running title: A cgMLST scheme for *Staphylococcus epidermidis* typing**

5 **Author names**

6 Anne Jamet^{1,2*}, Julien Guglielmini⁴, Bryan Brancotte⁴, Mathieu Coureuil¹, Daniel Euphrasie¹,
7 Julie Meyer¹, Johanna Roux¹, Jean-Philippe Barnier^{1,2}, Emmanuelle Bille^{1,2}, Agnès Ferroni²,
8 Jean-François Magny⁵, Christine Bôle-Feysot⁶, Alain Charbit¹, Xavier Nassif^{1,2}, Sylvain
9 Brisse^{3*}

10 **Affiliation**

11 ¹ Université de Paris, INSERM U1151 - CNRS UMR 8253, Institut Necker-Enfants Malades.
12 Team: Pathogenesis of Systemic Infections, Paris, France

13 ² Department of Clinical Microbiology, Necker Enfants-Malades Hospital, AP-HP.Centre
14 Université de Paris, Paris, France

15 ³ Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens Unit, 75015, Paris,
16 France

17 ⁴ Institut Pasteur, Bioinformatics and Biostatistics Hub, C3BI, USR 3756 IP CNRS, Paris,
18 France.

19 ⁵ Neonatal Intensive Care Unit, Necker Enfants-Malades Hospital, AP-HP.Centre Université
20 de Paris, Paris, France

21 ⁶ INSERM UMR 1163, Genomics platform, Université de Paris, Imagine Institute, Paris,
22 France; Genomic Core Facility, Imagine Institute-Structure Fédérative de Recherche Necker,
23 INSERM U1163, Paris, France

24
25
26 ***Corresponding authors**

27 Sylvain Brisse, Anne Jamet
28 sylvain.brisse@pasteur.fr ; anne.jamet@inserm.fr

29
30 **Keyword**

31 *Staphylococcus epidermidis*; cgMLST ; typing

32
33 **Abstract (max 250 words)**

34 *Staphylococcus epidermidis* is a pathogen emerging worldwide as a leading cause of health-
35 care associated infections. A standardized high-resolution typing method to document
36 transmission and dissemination of multidrug-resistant *S. epidermidis* strains is needed. Our
37 aim was to provide a core genome multilocus sequence typing (cgMLST) typing scheme for
38 *S. epidermidis* to improve the international surveillance of *S. epidermidis*.

39 We defined a cgMLST scheme based on 699 core genes and used it to investigate the
40 population structure of the species, and the genetic relatedness of isolates recovered from
41 infants hospitalized in several wards of a French hospital. Our results show the long-lasting
42 endemic persistence of *S. epidermidis* clones within and across wards of hospitals and
43 demonstrate the ability of our cgMLST approach to identify and track these clones.

44 We made the scheme publicly available through the Institut Pasteur BIGSdb server
45 (<http://bigsdb.pasteur.fr/epidermidis/>). This tool should enable international harmonization of
46 the epidemiological surveillance of multidrug-resistant *S. epidermidis* clones. By comparing
47 gene distribution among infection and commensal isolates, we also confirmed the association
48 of the *mecA* locus with infection isolates and of the *fdh* gene with commensal isolates.

49

50

51

52

53

54 **Introduction**

55 *Staphylococcus epidermidis*, a member of coagulase-negative staphylococci, is an
56 abundant and ubiquitous resident of the human skin and mucosal microbiota (1). The most
57 frequently isolated microorganisms in intensive care units (ICU)-acquired bloodstream
58 infections are coagulase-negative staphylococci, among which *S. epidermidis* is prominent
59 (2). *S. epidermidis* infections are almost exclusively associated with health care and occur
60 primarily in patients with implanted medical devices, such as intravascular devices or
61 orthopaedic and cardiac implants. In addition, neonates and immunocompromised patients
62 represent a population at particular risk for nosocomial infections (3). The ability of *S.*
63 *epidermidis* to form biofilms contributes to the chronicity and difficulties to treat these
64 infections (3, 4).

65 The proportion of MRSE (methicillin resistant *S. epidermidis*) among clinical isolates is
66 greater than 70% (5), leading to the probabilistic prescription of vancomycin in patients with
67 suspected *S. epidermidis* infections. Resistance to linezolid is rare but emerging (5, 6). The
68 emergence and global spread of multidrug-resistant *S. epidermidis* clones causing invasive
69 diseases has been largely documented (7, 8). In addition, the transmission of *S. epidermidis*
70 strains among patients and between patients and medical staff, mostly in neonatal intensive
71 care units, is possible (3, 9-11). However, the significance of *S. epidermidis* transmission is
72 likely to be underestimated, and the identification of subtypes that could represent higher risks
73 for transmission or pathogenesis, remains difficult using available subtyping methods.

74 The classification of *S. epidermidis* isolates has so far largely relied on Multilocus
75 Sequence Typing (MLST) based on the sequencing of 7 conserved loci (12). This typing
76 method allows the assignment of a Sequence Type (ST) to each unique allelic profile. STs can
77 be grouped into clonal complexes (CCs) with eBURST algorithm (13). While an extensive
78 strain-level diversity has been observed between healthy individuals (1), only a few health-
79 care associated clones are found in hospital environments worldwide (4, 14). Indeed, more
80 than 50% of the clinical strains and less than 10% of the community strains belong to a clonal
81 complex encompassing ST2 and ST5 (4, 5, 14). The global dissemination of these two
82 genotypes suggests an adaptation to the hospital environment. ST2 strains are almost all
83 resistant to methicillin (15, 16). Similar to the closely related *S. aureus* species, the *mecA*
84 gene confers resistance to methicillin and is found within the SCCmec (staphylococcal
85 cassette chromosome *mec*) genomic island (17, 18).

86 The high-resolution whole-genome sequencing (WGS) approach has been used to study *S.*

87 *epidermidis* linezolid-resistant isolates epidemiology (6, 19). A major insight from WGS-
88 based epidemiological studies is the description of long-lasting endemic persistence of *S.*
89 *epidermidis* clones across wards of hospitals (20, 21).

90 The population structure of *S. epidermidis* is divided into three main phylogenetic lineages
91 (17, 22), which encompass six previously defined clonal groups (23, 24). These subsets of *S.*
92 *epidermidis* show distinct associations with clinical infections and carriage. Besides, they
93 have distinctive genetic and phenotypic features including association with biofilm synthesis
94 genes, SSCmec and antimicrobial resistance. However, both asymptomatic carriage and
95 invasive isolates are found in all major phylogenetic groups (22), and the epidemiological
96 relationships of carriage and invasive isolates need to be deciphered at microevolutionary
97 timescales.

98 Genotyping of bacterial isolates is best performed with WGS for molecular
99 epidemiological studies (25-28). Core-genome MLST (cgMLST) provides the high resolution
100 of genome-wide methods while maintaining the benefits of standardization, reproducibility
101 and portability of classical 7-gene MLST (28, 29). While infections with multidrug resistant
102 *S. epidermidis* strains are an increasingly important issue in hospitals worldwide, an effective
103 tool for international strain comparison is still lacking.

104 The main aims of this study were i) To define a cgMLST scheme for *S. epidermidis*
105 genotyping; ii) To provide a publicly accessible international database for *S. epidermidis*
106 genomic typing; and (iii) To investigate the genetic relatedness of isolates recovered from
107 infants hospitalized in several wards of a French hospital, including a Neonatal Intensive Care
108 Unit (NICU). We also explored the population structure of *S. epidermidis* and the association
109 of previously reported marker genes and lineages.

110 **Methods**

111 **Routine procedures for bacterial isolation from clinical samples**

112 The 25 isolates sequenced here were recovered from one healthy adult and from cultures
113 performed routinely on clinical samples in the clinical microbiology laboratory at Necker-
114 Enfants Malades Hospital (Paris, France). The strains were grown on Columbia agar
115 supplemented with 5% horse blood and incubated 24-48 h at 37°C (INCUCCELL incubator,
116 MMM group, Germany). Coagulase-negative staphylococci isolates were identified at the
117 species level by MALDI-TOF MS (Bruker Daltonics, Bremen, Germany).

118

119 **Core genome MLST (cgMLST) scheme definition**

120 For the determination of the cgMLST scheme, 5 closed *S. epidermidis* genomes and 299 draft
121 genome assemblies available at NCBI GenBank (on the 22th of December 2015) were
122 downloaded. We also downloaded 83 draft genome assemblies deposited by Meric et al. (17)
123 in the Dryad repository. All used assemblies are publicly available from the Institut Pasteur
124 instance of the BIGSdb database at <https://bigsdb.pasteur.fr/epidermidis/>.

125 Low quality assemblies were excluded based on the following criteria: assembled genome
126 size >150 % of the reference genome; organism identified as non *S. epidermidis* as
127 determined using Kraken v0.10.5 (30) and Average Nucleotide Identity < 95% compared to
128 ATCC 12228 strain reference (GCF_000007645) (31). This resulted in the selection of 371
129 genomes for core genes definition. **Supplementary table S1** provides a detailed description
130 of the strain collection used for the establishment of the cgMLST scheme (e.g., date of
131 isolation, geographic origin and source).

132 From this set we inferred the species core genome using the CoreGeneBuilder pipeline default
133 settings (26) and *S. epidermidis* ATCC 12228 strain (GCF_000007645) as a reference. The
134 pipeline's first step relies on the eCAMBer software (32), which consists of a *de novo*
135 annotation of the genomes (except the reference) using Prodigal (33) and the harmonization
136 of the positions of the stop and start codons. In the next step, the core genome is inferred with
137 a bidirectional best hits (BBH) approach as previously described by Touchon et al. (34). We
138 used CoreGeneBuilder default settings. A gene was considered as part of the core genome if it
139 was found in at least 95% of the genomes under study. This resulted in an initial core genome
140 containing 1623 genes. We then filtered out some genes based on the following criteria. (i)
141 First, we removed 14 potential paralogous loci. The rationale behind this is that the presence
142 of paralogs inside a typing scheme can lead to ambiguities, as a candidate gene might be
143 attributable to two different core gene loci. To detect those potential paralogs, we compared
144 each allele of each locus against all the alleles of all the other loci using the software BLAT
145 (35). If a single hit was found between two different loci (more than 70% protein identity
146 between two alleles), we removed both loci. (ii) Second, we removed 840 loci whose length
147 varies among alleles, which is useful in reducing ambiguities during the genotyping process.
148 (iii) We also removed 10 loci for which alleles contained ambiguous characters. (iv) Last, we
149 removed 582 loci with variation within the start and stop codons; the rationale is that allele
150 calling relies on BLASTn, which may not produce full-length alignments when the start
151 or stop codons differ, resulting in incomplete alleles. Note that these sets of filtered loci can
152 overlap. After applying these four steps, 699 core genes were retained and constitute our
153 cgMLST scheme.

154

155 **cgMLST allele definitions and tagging of publicly available genomes using BIGSdb**

156 The BIGSdb server is designed to host separately a database of allele sequence definitions and
157 a database of genomic assemblies with attached source information (36). Newly imported
158 assemblies were scanned using BIGSdb to designate alleles of each locus. A locus was tagged
159 when its sequence had > 90% nucleotide sequence identity over at least 90% of the alignment
160 length with a reference allele. In case of exact match with an allele found in the database, the
161 locus was designated by the allele number. In case of partial match, if the locus corresponded
162 to a protein-coding sequence (beginning with a start codon and ending with a stop codon), the
163 allele was added to the allele database. Then, a subsequent scan allowed tagging isolates with
164 novel alleles. A locus was considered missing when there was no match with alleles in the
165 database.

166 To evaluate the conservation of the chosen genes in an independent sample of the
167 *S. epidermidis* population, we analyzed 289 additional genome assemblies from NCBI
168 GenBank and from Harris *et al.* (4) that were not yet available at the time of scheme
169 definition. The validation set is described in **Supplementary table S1**.

170 All assemblies, allele sequences designations and profiles are available at
171 <http://bigsdbs.pasteur.fr/epidermidis/>.

172

173 **Assessment of the population structure of *S. epidermidis***

174 The population structure of *S. epidermidis* was assessed based on 703 available genomes. To
175 obtain the alignment required for the phylogenetic analysis, we extracted, for each genome
176 assembly, the allelic sequences of the 699 loci. For each locus the alleles sequences were
177 translated to proteins, aligned using MAFFT (37) and the G-INS-i algorithm, and back-
178 translated to nucleotide sequences, thus resulting in codon alignments. We finally
179 concatenated the 699 alignments into one supermatrix, containing 703 sequences (isolates)
180 and 539 151 sites (the nucleotide positions of missing alleles were replaced by gaps in the
181 concatenate of alignments of individual loci). A phylogenetic tree was obtained with the
182 maximum likelihood (ML) criterion after purging recombination using Gubbins v2.2.0 (38)
183 with default parameters resulting in an alignment of 31,385 sites. RAxML v8.2.8 (39) was
184 used for the phylogenetic inference under the GTRCAT model (final tree optimized under the
185 GTRGAMMA model). The tree was drawn using iTOL (40).

186 The same methodology was applied to compare clusters obtained based on cgMLST allelic
187 profiles with those observed by phylogenetic analysis of the concatenated allele sequences of
188 the 699 core loci.

189

190 **Patients and isolates**

191 To evaluate the ability of cgMLST to identify potential transmission events within hospital
192 settings, we selected 25 isolates including 24 clinical isolates from patients hospitalized in
193 several wards between November 2014 and August 2016.

194 The 25 isolates were recovered from 21 hospitalized individuals (16 hospitalized neonates and
195 5 hospitalized children) and one non-hospitalized healthy adult. All neonates have stayed in a
196 NICU, 3 children have stayed in a gastroenterology ward and 2 children have stayed in an
197 orthopedic ward. One isolate was sequenced per patient except for 3 patients who had 2
198 isolates sequenced (patients 3, 7 and 21). Information on patients and sources of isolates is
199 summarized in **Table 1**.

200 The study of *S. epidermidis* clinical isolates collected from the Necker-Enfants Malades
201 Hospital was approved by the ethics review board of our institution and registered with
202 clinicaltrials.gov Identifier#NCT03374371. All experiments were performed in accordance
203 with the guidelines and regulations described by the Declaration of Helsinki and the low
204 Huriet-Serusclet on human research ethics and informed consent was obtained for all
205 participating subjects.

206

207 **Whole-genome sequencing and assembly**

208 DNA extractions were performed on pure culture using DNeasy Blood and Tissue Kit
209 (Qiagen) following manufacturer's instructions with addition of lysostaphin (Sigma) to the
210 Gram-positive lysis buffer (final concentration 20 µg/mL).

211 Whole-genome shotgun libraries sequenced on an Illumina MiSeq sequencing platform (2 x
212 150 bp paired-end reads) were prepared with the Nextera XT Kit (Illumina). Whole-genome
213 shotgun libraries sequenced on an Illumina HiSeq-2500 sequencing platform (2 x 130 bp
214 paired-end reads) were prepared with the Ovation Ultralow V2 kit (NuGEN).

215 Reads were trimmed using Trimmomatic v0.36 (41) and assembled using SPAdes v3.9.0 (42).
216 *De novo* assemblies quality was assessed with QUASt (43). *In silico* MLST was performed
217 on the *de novo* assemblies with MLST v2.6 (<https://github.com/tseemann/mlst>). *De novo*
218 assemblies were imported and scanned in BIGSdb.

219

220 **Cluster analysis of clinical isolates using BIGSdb**

221 To validate the ability of cgMLST to cluster related isolates, we first used sequence data from
222 a recently published study of linezolid resistant isolates (6). Raw Illumina reads (accession
223 number SRP039360) were downloaded from NCBI Sequence Read Archive (SRA) and
224 processed as newly sequenced genomes. Assemblies were imported and scanned in BIGSdb
225 except isolate UCLA-Sa-LR-01, which was discarded due to poor quality of its *de novo*
226 assembly. We also tested our cgMLST scheme using the 25 newly sequenced clinical isolates
227 (**Table 1**). The PHYLOViZ tool (integrated as a plugin of BIGSdb) was used to generate
228 minimum spanning trees based on the cgMLST allelic profiles (44). Here, clusters were
229 defined as groups of genotypes showing ≤ 21 cgMLST allele mismatches with at least one
230 other member of the group.

231

232 **Data availability**

233 This Whole Genome Shotgun project has been deposited in GenBank under the Bioproject ID
234 PRJNA608162.

235 All assemblies, sequences and their allelic designations and cgMLST profiles are available
236 from the Institut Pasteur BIGSdb database access web page at
237 <https://bigsd.bpasteur.fr/epidermidis/>.

238

239 **Results**

240 **Definition and evaluation of a cgMLST scheme for *S. epidermidis* using public genomes**

241 Using 371 publically available genomes, we developed a cgMLST scheme consisting of 699
242 core-genes, which includes 30% of the 2,348 CDS of *S. epidermidis* ATCC 12228 reference
243 genome. As shown in **Supplementary figure S1**, the distribution along the *S. epidermidis*
244 genome of the 699 cgMLST genes was even.

245 To evaluate the conservation of gene loci of the newly defined cgMLST scheme, a
246 validation set of 289 additional publically available genome assemblies was analyzed (NCBI
247 genomes, last accessed on Sept. 27th, 2017). These additional genomes were made public after
248 the definition of our cgMLST scheme and thus represent a non-overlapping strain collection.
249 These genomes had 97.5% tagged loci on average, showing that the cgMLST scheme loci are
250 highly conserved, and can therefore be expected to be mostly present in future sequenced
251 genomes. **Supplementary figure S2** shows the distribution of the number of missing loci
252 among the 699 genes of the scheme.

253 The cgMLST genotyping system is publicly available through the Institut Pasteur BIGSdb
254 server (<https://bigsd.bpasteur.fr/epidermidis/>).

255

256 **Population structure of *S. epidermidis* genomes**

257 We next used 703 genomes to assess the phylogenetic structure of *S. epidermidis* based on
258 cgMLST gene sequences (**Supplementary Figure 1**), which confirmed that the population is
259 divided into 3 main clades (A-C) as previously reported (17). Most of the isolates belong to
260 clade A, which encompasses clonal groups 5 and 6 associated with the two major nosocomial
261 clones, ST2 and ST5 respectively (23, 24).

262 The **supplementary figure S3** shows the concordance with clades A, B and C as defined
263 previously by Méric and colleagues (17), with the clonal groups defined previously (23, 24),
264 and with classical 7-gene MLST.

265 In addition to the core genes, we scanned the assemblies for the presence of accessory genes
266 associated with the SCCmec locus (*mecA*, *ccrA* and *ccrB*) and the *ica* genes associated with
267 biofilm formation capacity through polymeric N-acetyl-glucosamine (PNAG) synthesis. In
268 addition, we searched for the putative commensal marker *fdh* encoding a formate
269 dehydrogenase (14).

270 The distribution of SCCmec associated genes among infection and commensal isolates
271 showed a positive association between infection strains and the SCCmec locus (Fisher <1e-3)
272 (**Figure 1**). The search of the co-occurrence of the *icaAD* genes, commonly used as a proxy to
273 evaluate the biofilm formation potential of isolates (45), yielded no significant association
274 (**Figure 1**). In contrast, a positive association between commensal isolates compared to
275 infection isolates was observed for the putative commensal marker gene *fdh* (Fisher <0.005)
276 (**Figure 1**).

277

278 **Application of the *S. epidermidis* cgMLST scheme to hospital linezolid-resistant isolates**

279 To evaluate the ability of cgMLST to cluster *S. epidermidis* isolates, we analyzed the
280 relationships among isolates from a previous study of linezolid resistant isolates (6).

281 All genomes had > 93% of tagged loci (95.5% on average). Cluster analysis based on
282 cgMLST allelic profiles showed 4 clusters and 1 outgroup isolate (**Figure 2**). The same
283 clusters were found after phylogenetic analysis of the concatenated sequences of core genes
284 (**Figure 2**). These results were concordant with the ones reported by Tewhey *et al.* based on
285 the alignments of 2,093 core genes (6). However, there was no description of potential

286 epidemiological links between the isolates of this study, which were compared only because
287 of their common phenotypic linezolid resistance.

288

289 **Application to *S. epidermidis* epidemiology in a French hospital**

290 To assess the ability of cgMLST to cluster *S. epidermidis* isolates with defined
291 epidemiological sources, we next sequenced 24 clinical isolates recovered from several wards
292 of a pediatric French hospital and an isolate from a healthy adult (total of 25 sequenced
293 isolates). The mean number of contigs was 211 (range, 137 to 410) (**Table S2**). After
294 cgMLST genotyping of the genomic assemblies, the mean percentage of tagged loci was
295 97.6% (range, 96.1 to 100%) (**Table S2**). As expected, most isolates recovered from
296 hospitalized patients (20 out of 24) belonged to ST2 (n = 12) and ST5 (n = 8) (**Table 1**).

297 Nineteen isolates were grouped into 6 clusters (≤ 21 allele differences), whereas there were 6
298 ungrouped isolates (**Figure 3**). Five clusters were observed among isolates from the NICU
299 during one year (span time of clusters ranging from 49 to 354 days), whereas another cluster
300 contained isolates from the orthopedic and the gastroenterology wards. The maximum number
301 of differing alleles among isolates of the same cluster was 21, whereas the minimum number
302 of differing alleles between isolates of distinct clusters was 39 (**Figure 3**). For the 3 pairs of
303 isolates recovered from 3 individual patients, collected between 7 to 38 days apart, we
304 observed a maximum of 7 differences. The timeline shown in **Figure 4** illustrates that isolates
305 belonging to the same cluster were sometimes found in patients hospitalized in distinct wards
306 (*i.e.*, cluster IV in patients hospitalized in gastroenterology and orthopedic units) and in
307 patients hospitalized up to 1-year apart (*i.e.*, cluster I for patients 3, 16 and 18).

308

309 **Discussion**

310 *S. epidermidis* is a leading cause of bacteremia and device-related infections and can cause
311 severe infections in immunocompromised patients and neonates. *S. epidermidis* infections are
312 mostly caused by strains belonging to multidrug-resistant nosocomial clones ST2 and ST5 (5,
313 10, 19, 46). Our results corroborate the predominance of ST2 and ST5 among hospitalized
314 patients. Although the 7-gene MLST has played an important role to recognize the worldwide
315 emergence and nosocomial endemic persistence of multidrug-resistant clones, there is a need
316 for higher resolution genotyping, allowing strain comparison within endemic clones.
317 Moreover, standardization is paramount for worldwide tracking of strain dissemination and

318 recognizing subtypes with medically important properties. We therefore explored a core
319 genome MLST (cgMLST) approach, which provides a high resolution and has the potential
320 for a universal typing method both for epidemiological investigations and for population
321 biology studies. Standardized cgMLST schemes are available for many pathogens including
322 *Klebsiella pneumoniae*, *Campylobacter jejuni*, *Campylobacter coli*, *Staphylococcus aureus*
323 and *Listeria monocytogenes* (28, 47-49). However, a cgMLST scheme has been lacking until
324 now for *S. epidermidis*.

325 The potential of the developed *S. epidermidis* cgMLST scheme to delineate closely related
326 isolates was investigated by comparing genomes obtained from children hospitalized in
327 several wards of the same hospital. Our analysis showed that the cgMLST scheme of 699
328 genes had an adequate discriminatory power to delineate distinct clusters of very closely
329 related isolates within the same ST. The maximum allelic distance among epidemiologically
330 related isolates (2 isolates recovered from 2 patients hospitalized several months apart in the
331 orthopedic ward) was 21, and we chose this cutoff to define clusters. This relatively high
332 number of differences is likely to reflect the fact that clones are able to persist several years in
333 hospitals where they diversify and are likely transferred to patients on multiple occasions (7,
334 20, 21). For instance, Dortet *et al.* investigated a long-lasting outbreak spanning 2 years
335 caused by linezolid-resistant *S. epidermidis* clones and report 22 ST2 isolates with 13 SNPs
336 along their core genome (19). Our findings also show that several clones can co-exist and
337 persist over months in a NICU. This is concordant with previous observations showing that
338 patients acquire genetically related strains belonging to multidrug-resistant endemic clones
339 that persist in clinical settings (10, 50).

340 In addition, we leveraged the ability of the BIGSdb database tool to analyse the presence of
341 genes of medical interest to compare the distribution of several putative marker genes
342 associated with infection or carriage. We confirmed a positive association between infection
343 strains and the SCCmec locus (22) and a positive association between commensal isolates and
344 the *fdh* gene (14). The co-occurrence of *icaAD* genes has been proposed as a marker for
345 biofilm formation ability (45) linked with isolate virulence. However, the presence of the *ica*
346 locus does not always correlate with biofilm production *in vitro* (45). For instance in a study
347 of 200 *S. epidermidis* isolates from hospitalized patients in Germany, among 124 *icaAD*-
348 positive isolates, only 49% were able to produce biofilm *in vitro*, whereas only 50% of *in*
349 *vitro* biofilm-producing isolates were *ica*-positive (51). The weak correlation between *ica*
350 presence and biofilm formation ability is corroborated by Du *et al.* who reported biofilm
351 production in less than 60% of their *icaA*-positive isolates (52).

352 In another study, when nosocomial isolates associated with infection or contamination were
353 compared only *mecA*, but neither *icaA* nor *fdh*, was differentially found among infection and
354 contamination isolates (24).

355

356 In conclusion, we developed a cgMLST scheme for WGS-based typing of *S. epidermidis*,
357 integrated it in a publically available BIGSdb web application platform and illustrated its
358 application to the molecular epidemiology of *S. epidermidis* in a French university hospital.

359 We hope that the novel genotyping tool will contribute to allowing reproducible comparisons
360 among clinical laboratories for the epidemiological surveillance and biological studies of *S.*
361 *epidermidis* clones causing infections.

362

363

364

365

366

367 **Author statements**

368 **Conflicts of interest**

369 The authors declare that there are no conflicts of interest

370 **Funding information:** Contrat de Recherche Clinique, French Ministry of Health
371 (CRC17040); the sponsor was Assistance Publique des Hôpitaux de Paris (APHP)
372 (Département de la Recherche Clinique et du Développement).

373 **Ethical approval:** The study of *S. epidermidis* clinical isolates collected from the Necker-
374 Enfants Malades Hospital has been approved by the ethics review board of our institution and
375 registered with clinicaltrials.gov Identifier#NCT03374371. All experiments were performed
376 in accordance with the guidelines and regulations described by the Declaration of Helsinki
377 and the law Huriot-Serusclet on human research ethics and informed consent was obtained for
378 all participating subjects.

379

380 **Acknowledgements**

381 The authors want to thank Pr C. Bréchet for having stimulated the collaboration between the
382 Necker Hospital and Institut Pasteur in the frame of the Pasteur International biobanking
383 network (PIBnet) programme. We thank Bryan Brancotte for maintenance of the Institut
384 Pasteur web application.

385 **References**

386

- 387 1. Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, Segre JA. 2014.
388 Biogeography and individuality shape function in the human skin metagenome.
389 Nature 514:59-64. <https://doi.org/10.1038/nature13786>
- 390 2. Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. 2004.
391 Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a
392 prospective nationwide surveillance study. Clin Infect Dis 39:309-17.
393 <https://doi.org/10.1086/421946>
- 394 3. Becker K, Heilmann C, Peters G. 2014. Coagulase-negative staphylococci. Clin
395 Microbiol Rev 27:870-926. <https://doi.org/10.1128/CMR.00109-13>
- 396 4. Harris LG, Murray S, Pascoe B, Bray J, Meric G, Magerios L, Wilkinson TS, Jeeves
397 R, Rohde H, Schwarz S, de Lencastre H, Miragaia M, Rolo J, Bowden R, Jolley KA,
398 Maiden MC, Mack D, Sheppard SK. 2016. Biofilm Morphotypes and Population
399 Structure among Staphylococcus epidermidis from Commensal and Clinical Samples.
400 PLoS One 11:e0151240. <https://doi.org/10.1371/journal.pone.0151240>

- 401 5. Mendes RE, Deshpande LM, Costello AJ, Farrell DJ. 2012. Molecular epidemiology
402 of *Staphylococcus epidermidis* clinical isolates from U.S. hospitals. *Antimicrob*
403 *Agents Chemother* 56:4656-61. <https://doi.org/10.1128/AAC.00279-12>
- 404 6. Tewhey R, Gu B, Kelesidis T, Charlton C, Bobenchik A, Hindler J, Schork NJ,
405 Humphries RM. 2014. Mechanisms of linezolid resistance among coagulase-negative
406 staphylococci determined by whole-genome sequencing. *MBio* 5:e00894-14.
407 <https://doi.org/10.1128/mBio.00894-14>
- 408 7. Lee JYH, Monk IR, Goncalves da Silva A, Seemann T, Chua KYL, Kearns A, Hill R,
409 Woodford N, Bartels MD, Strommenger B, Laurent F, Dodemont M, Deplano A, Patel
410 R, Larsen AR, Korman TM, Stinear TP, Howden BP. 2018. Global spread of three
411 multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat Microbiol* 3:1175-
412 1185. <https://doi.org/10.1038/s41564-018-0230-7>
- 413 8. Li X, Arias CA, Aitken SL, Galloway Pena J, Panesso D, Chang M, Diaz L, Rios R,
414 Numan Y, Ghaoui S, DebRoy S, Bhatti MM, Simmons DE, Raad I, Hachem R, Folan
415 SA, Sahasarabhojane P, Kalia A, Shelburne SA. 2018. Clonal Emergence of Invasive
416 Multidrug-Resistant *Staphylococcus epidermidis* Deconvoluted via a Combination of
417 Whole-Genome Sequencing and Microbiome Analyses. *Clin Infect Dis* 67:398-406.
418 <https://doi.org/10.1093/cid/ciy089>
- 419 9. Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, Shendure J,
420 Salipante SJ. 2015. A Year of Infection in the Intensive Care Unit: Prospective Whole
421 Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and
422 Novel Microbiota. *PLoS Genet* 11:e1005413.
423 <https://doi.org/10.1371/journal.pgen.1005413>
- 424 10. Soeorg H, Huik K, Parm U, Ilmoja ML, Metsvaht T, Lutsar I. 2017. Molecular
425 epidemiology of *Staphylococcus epidermidis* in neonatal intensive care units. *APMIS*
426 125:63-73. <https://doi.org/10.1111/apm.12637>
- 427 11. Milisavljevic V, Wu F, Cimmotti J, Haas J, Della-Latta P, Larson E, Saiman L. 2005.
428 Genetic relatedness of *Staphylococcus epidermidis* from infected infants and staff in
429 the neonatal intensive care unit. *Am J Infect Control* 33:341-7.
430 <https://doi.org/10.1016/j.ajic.2005.02.003>
- 431 12. Thomas JC, Vargas MR, Miragaia M, Peacock SJ, Archer GL, Enright MC. 2007.
432 Improved multilocus sequence typing scheme for *Staphylococcus epidermidis*. *J Clin*
433 *Microbiol* 45:616-9. <https://doi.org/10.1128/JCM.01934-06>
- 434 13. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring
435 patterns of evolutionary descent among clusters of related bacterial genotypes from
436 multilocus sequence typing data. *J Bacteriol* 186:1518-30.
- 437 14. Conlan S, Mijares LA, Program NCS, Becker J, Blakesley RW, Bouffard GG, Brooks
438 S, Coleman H, Gupta J, Gurson N, Park M, Schmidt B, Thomas PJ, Otto M, Kong
439 HH, Murray PR, Segre JA. 2012. *Staphylococcus epidermidis* pan-genome sequence
440 analysis reveals diversity of skin commensal and hospital infection-associated isolates.
441 *Genome Biol* 13:R64. <https://doi.org/10.1186/gb-2012-13-7-r64>
- 442 15. Argudin MA, Vanderhaeghen W, Vandendriessche S, Vandecandelaere I, Denis O,
443 Coenye T, Butaye P. 2015. Biofilm formation of *ica* operon-positive *Staphylococcus*
444 *epidermidis* from different sources. *APMIS* 123:1081-9.
445 <https://doi.org/10.1111/apm.12472>
- 446 16. Lee JY, Monk IR, Pidot SJ, Singh S, Chua KY, Seemann T, Stinear TP, Howden BP.
447 2016. Functional analysis of the first complete genome sequence of a multidrug
448 resistant sequence type 2 *Staphylococcus epidermidis*. *Microbial Genomics*.
- 449 17. Meric G, Miragaia M, de Been M, Yahara K, Pascoe B, Mageiros L, Mikhail J, Harris
450 LG, Wilkinson TS, Rolo J, Lamble S, Bray JE, Jolley KA, Hanage WP, Bowden R,

- 451 Maiden MC, Mack D, de Lencastre H, Feil EJ, Corander J, Sheppard SK. 2015.
452 Ecological Overlap and Horizontal Gene Transfer in *Staphylococcus aureus* and
453 *Staphylococcus epidermidis*. *Genome Biol Evol* 7:1313-28.
454 <https://doi.org/10.1093/gbe/evv066>
- 455 18. Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT,
456 Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli
457 SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G,
458 Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM. 2005.
459 Insights on evolution of virulence and resistance from the complete genome analysis
460 of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing
461 methicillin-resistant *Staphylococcus epidermidis* strain. *J Bacteriol* 187:2426-38.
462 <https://doi.org/10.1128/JB.187.7.2426-2438.2005>
- 463 19. Dortet L, Glaser P, Kassis-Chikhani N, Girlich D, Ichai P, Boudon M, Samuel D,
464 Creton E, Imanci D, Bonnin R, Fortineau N, Naas T. 2018. Long-lasting successful
465 dissemination of resistance to oxazolidinones in MDR *Staphylococcus epidermidis*
466 clinical isolates in a tertiary care hospital in France. *J Antimicrob Chemother* 73:41-
467 51. <https://doi.org/10.1093/jac/dkx370>
- 468 20. Chong J, Quach C, Blanchard AC, Poliquin PG, Golding GR, Laferriere C, Levesque
469 S. 2016. Molecular Epidemiology of a Vancomycin-Intermediate Heteroresistant
470 *Staphylococcus epidermidis* Outbreak in a Neonatal Intensive Care Unit. *Antimicrob*
471 *Agents Chemother* 60:5673-81. <https://doi.org/10.1128/AAC.00726-16>
- 472 21. Rodriguez-Lucas C, Rodicio MR, Camara J, Dominguez MA, Alaguero M, Fernandez
473 J. 2019. Long-term endemic situation caused by a linezolid- and methicillin-resistant
474 clone of *Staphylococcus epidermidis* in a tertiary hospital. *J Hosp Infect*
475 doi:10.1016/j.jhin.2019.10.013. <https://doi.org/10.1016/j.jhin.2019.10.013>
- 476 22. Meric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, Kittiwat N, Tadee P,
477 Post V, Lamble S, Bowden R, Bray JE, Morgenstern M, Jolley KA, Maiden MCJ, Feil
478 EJ, Didelot X, Miragaia M, de Lencastre H, Moriarty TF, Rohde H, Massey R, Mack
479 D, Corander J, Sheppard SK. 2018. Disease-associated genotypes of the commensal
480 skin bacterium *Staphylococcus epidermidis*. *Nat Commun* 9:5034.
481 <https://doi.org/10.1038/s41467-018-07368-7>
- 482 23. Thomas JC, Zhang L, Robinson DA. 2014. Differing lifestyles of *Staphylococcus*
483 *epidermidis* as revealed through Bayesian clustering of multilocus sequence types.
484 *Infect Genet Evol* 22:257-64. <https://doi.org/10.1016/j.meegid.2013.06.020>
- 485 24. Tolo I, Thomas JC, Fischer RSB, Brown EL, Gray BM, Robinson DA. 2016. Do
486 *Staphylococcus epidermidis* Genetic Clusters Predict Isolation Sources? *J Clin*
487 *Microbiol* 54:1711-1719. <https://doi.org/10.1128/JCM.03345-15>
- 488 25. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT.
489 2015. Application of whole-genome sequencing for bacterial strain typing in
490 molecular epidemiology. *J Clin Microbiol* 53:1072-9.
491 <https://doi.org/10.1128/JCM.03385-14>
- 492 26. Guglielmini J, Bourhy P, Schiettekatte O, Zinini F, Brisse S, Picardeau M. 2019.
493 Genus-wide *Leptospira* core genome multilocus sequence typing for strain taxonomy
494 and global surveillance. *PLoS Negl Trop Dis* 13:e0007374.
495 <https://doi.org/10.1371/journal.pntd.0007374>
- 496 27. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D,
497 Mellmann A. 2015. Defining and Evaluating a Core Genome Multilocus Sequence
498 Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria*
499 *monocytogenes*. *J Clin Microbiol* 53:2869-76. <https://doi.org/10.1128/JCM.01193-15>

- 500 28. Cody AJ, Bray JE, Jolley KA, McCarthy ND, Maiden MCJ. 2017. Core Genome
501 Multilocus Sequence Typing Scheme for Stable, Comparative Analyses of
502 *Campylobacter jejuni* and *C. coli* Human Disease Isolates. *J Clin Microbiol* 55:2086-
503 2097. <https://doi.org/10.1128/JCM.00080-17>
- 504 29. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA,
505 McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial
506 genomics. *Nat Rev Microbiol* 11:728-36. <https://doi.org/10.1038/nrmicro3093>
- 507 30. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification
508 using exact alignments. *Genome Biol* 15:R46. [https://doi.org/10.1186/gb-2014-15-3-
509 r46](https://doi.org/10.1186/gb-2014-15-3-r46)
- 510 31. Haubold B, Klotzl F, Pfaffelhuber P. 2015. andi: fast and accurate estimation of
511 evolutionary distances between closely related genomes. *Bioinformatics* 31:1169-75.
512 <https://doi.org/10.1093/bioinformatics/btu815>
- 513 32. Wozniak M, Wong L, Tiuryn J. 2014. eCAMBer: efficient support for large-scale
514 comparative analysis of multiple bacterial strains. *BMC Bioinformatics* 15:65.
515 <https://doi.org/10.1186/1471-2105-15-65>
- 516 33. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
517 prokaryotic gene recognition and translation initiation site identification. *BMC*
518 *Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- 519 34. Touchon M, Hoede C, Tenailon O, Barbe V, Baeriswyl S, Bidet P, Bingen E,
520 Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller
521 S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles
522 AM, Johnson J, Le Bouguéne C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic
523 I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Turret J,
524 Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E. 2009. Organised
525 Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive
526 Paths. *PLoS Genetics* 5:e1000344. <https://doi.org/10.1371/journal.pgen.1000344>
- 527 35. Kent WJ. 2002. BLAT--The BLAST-Like Alignment Tool. *Genome Research*
528 12:656-664. <https://doi.org/10.1101/gr.229202>
- 529 36. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome
530 variation at the population level. *BMC Bioinformatics* 11:595.
531 <https://doi.org/10.1186/1471-2105-11-595>
- 532 37. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
533 7: improvements in performance and usability. *Mol Biol Evol* 30:772-80.
534 <https://doi.org/10.1093/molbev/mst010>
- 535 38. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J,
536 Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant
537 bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15.
538 <https://doi.org/10.1093/nar/gku1196>
- 539 39. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
540 analysis of large phylogenies. *Bioinformatics* 30:1312-3.
541 <https://doi.org/10.1093/bioinformatics/btu033>
- 542 40. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
543 developments. *Nucleic Acids Res* 47:W256-W259. <https://doi.org/10.1093/nar/gkz239>
- 544 41. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
545 sequence data. *Bioinformatics* 30:2114-20.
546 <https://doi.org/10.1093/bioinformatics/btu170>
- 547 42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
548 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
549 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and

- 550 its applications to single-cell sequencing. *J Comput Biol* 19:455-77.
551 <https://doi.org/10.1089/cmb.2012.0021>
- 552 43. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool
553 for genome assemblies. *Bioinformatics* 29:1072-5.
554 <https://doi.org/10.1093/bioinformatics/btt086>
- 555 44. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carrico JA, Vaz C. 2017.
556 PHYLOViZ 2.0: providing scalable data integration and visualization for multiple
557 phylogenetic inference methods. *Bioinformatics* 33:128-129.
558 <https://doi.org/10.1093/bioinformatics/btw582>
- 559 45. Cafiso V, Bertuccio T, Santagati M, Campanile F, Amicosante G, Perilli MG, Selan L,
560 Artini M, Nicoletti G, Stefani S. 2004. Presence of the ica operon in clinical isolates of
561 *Staphylococcus epidermidis* and its role in biofilm production. *Clin Microbiol Infect*
562 10:1081-8. <https://doi.org/10.1111/j.1469-0691.2004.01024.x>
- 563 46. Shelburne SA, Dib RW, Endres BT, Reitzel R, Li X, Kalia A, Sahasrabhojane P,
564 Chaftari AM, Hachem R, Vargas-Cruz NS, Jiang Y, Garey K, Fowler VG, Jr., Holland
565 TL, Gu J, Miller W, Sakurai A, Arias CA, Aitken SL, Greenberg DE, Kim J, Flores
566 AR, Raad I. 2020. Whole-genome sequencing of *Staphylococcus epidermidis*
567 bloodstream isolates from a prospective clinical trial reveals that complicated
568 bacteraemia is caused by a limited number of closely related sequence types. *Clin*
569 *Microbiol Infect* 26:646 e1-646 e8. <https://doi.org/10.1016/j.cmi.2019.10.008>
- 570 47. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS,
571 Garin B, Le Hello S, Arlet G, Nicolas-Chanoine MH, Decre D, Brisse S. 2014.
572 Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae*
573 clonal groups. *Emerg Infect Dis* 20:1812-20. <https://doi.org/10.3201/eid2011.140206>
- 574 48. Moura A, Tourdjman M, Leclercq A, Hamelin E, Laurent E, Fredriksen N, Van
575 Cauteren D, Bracq-Dieye H, Thouvenot P, Vales G, Tessaud-Rita N, Maury MM,
576 Alexandru A, Criscuolo A, Quevillon E, Donguy MP, Enouf V, de Valk H, Brisse S,
577 Lecuit M. 2017. Real-Time Whole-Genome Sequencing for Surveillance of *Listeria*
578 *monocytogenes*, France. *Emerg Infect Dis* 23:1462-1470.
579 <https://doi.org/10.3201/eid2309.170336>
- 580 49. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. 2014. Bacterial whole-
581 genome sequencing revisited: portable, scalable, and standardized analysis for typing
582 and detection of virulence and antibiotic resistance genes. *J Clin Microbiol* 52:2365-
583 70. <https://doi.org/10.1128/JCM.00262-14>
- 584 50. Krediet TG, Mascini EM, van Rooij E, Vlooswijk J, Paauw A, Gerards LJ, Fleer A.
585 2004. Molecular epidemiology of coagulase-negative staphylococci causing sepsis in a
586 neonatal intensive care unit over an 11-year period. *J Clin Microbiol* 42:992-5.
- 587 51. Mertens A, Ghebremedhin B. 2013. Genetic determinants and biofilm formation of
588 clinical *Staphylococcus epidermidis* isolates from blood cultures and indwelling
589 devices. *Eur J Microbiol Immunol (Bp)* 3:111-9.
590 <https://doi.org/10.1556/EuJMI.3.2013.2.4>
- 591 52. Du X, Zhu Y, Song Y, Li T, Luo T, Sun G, Yang C, Cao C, Lu Y, Li M. 2013.
592 Molecular analysis of *Staphylococcus epidermidis* strains isolated from community
593 and hospital environments in China. *PLoS One* 8:e62742.
594 <https://doi.org/10.1371/journal.pone.0062742>

595

596

597

598 **Figure legends**

599 **Figure 1. Distribution of gene presence and absence for *mecA*, *ccrAB*, *ica* and *fdh* loci**
600 **among 252 infection isolates and 194 commensal isolates from the BIGSdb database.**

601 *mecA* encodes penicillin-binding protein 2 (PBP2); *ccrA* and *ccrB* encode cassette
602 chromosome recombinase A and B respectively; *icaA* and *icaD* genes are involved in
603 polysaccharide intercellular adhesin (PIA) production; *fdh* encodes a formate dehydrogenase.
604 There was a positive association between infection isolates and *mecA*, *ccrA* and *ccrB* loci
605 compared to commensal isolates (** Fisher<0.001) and there was a positive association
606 between commensal isolates and *fdh* locus compared to infection isolates (* Fisher <0.005).

607

608

609

610 **Figure 2. Comparison of cluster analysis based on cgMLST allelic profiles with**
611 **phylogenetic analysis based on core genes alignments for 20 linezolid-resistant isolates.**

612 The minimum-spanning tree based on cgMLST allelic profiles is shown on the left side. The
613 tree was generated from allelic profiles of 699 core genes and visualised with PHYLOViZ.
614 Each node represents an isolate and is coloured according to the 7-loci ST. *In silico* MLST
615 performed on the *de novo* assemblies revealed that isolates UCLA-Sa-LR-14 (497) and
616 UCLA-Sa-LR-21 (503) had been erroneously assigned to ST24 and ST186-like respectively
617 in the original publication (6). We found that in fact both strains belonged to ST186. Black
618 numbers next to the nodes indicate the BIGSdb identifiers of each isolate. Red numbers on the
619 connecting lines indicate the number of core gene allele differences between adjacent nodes.
620 Link lengths are loosely related to the number of allelic differences.

621 The tree based on 699 core gene sequences alignments is shown on the right side. The tip
622 labels indicate the BIGSdb identifiers of each isolate.

623 Clusters reported by Tewhey *et al.* based on the alignments of 2,093 core genes (6) are
624 indicated by circles and letters (A, B, C and D).

625

626

627

628 **Figure 3. Comparison of clustering analyses based on cgMLST allelic profiles and on**
629 **core gene sequences alignment, for 25 *S. epidermidis* isolates sequenced in this study.**

630 The minimum-spanning tree based on cgMLST allelic profiles is shown on the left side. The
631 tree was generated from allelic profiles of 699 core genes. Each node represents an allelic

632 profile and corresponds to a single isolate except the node corresponding to cluster VI where
633 2 isolates have the same profile. Nodes are coloured according to the ward where patients
634 were hospitalised. Black numbers next to the nodes indicate the BIGSdb identifiers of each
635 isolate. Red numbers on the connecting lines indicate the number of core gene alleles
636 differences between adjacent nodes. Link lengths are not proportional to the number of allelic
637 differences. Clusters of closely related genotypes (≤ 21 allele differences) are surrounded by
638 circles and are numbered (I to VI).

639 The tree based on 699 core genes alignment is shown on the right side. The tip labels indicate
640 the BIGSdb identifiers and names of each isolate. Clusters are indicated by circles and are
641 numbered (I to VI) as in the minimum-spanning tree.

642

643

644 **Figure 4. Timeline of the 21 hospitalized patients from which 24 *S. epidermidis* isolates**
645 **were sequenced.**

646 The patient numbers indicated in the coloured boxes and the patient isolate numbers indicated
647 between brackets correspond to those in **Table 1**. The clusters to which isolates belong
648 correspond to those in Figure 3. The isolate sequence type (ST) is indicated on the right side
649 of the graph.

650