Functional genomics reveals extensive diversity in *Staphylococcus epidermidis* restriction modification systems compared to *Staphylococcus aureus*.

Authors: Jean YH Lee*[1], Glen P Carter[1,2,3], Sacha J Pidot[1], Romain Guérillot[1], Torsten Seemann[1,2,3], Anders Gonçalves da Silva[2,3], Timothy J Foster[4], Benjamin P Howden[#2,3,5], Timothy P Stinear[#1,3], Ian R Monk*[#1]

[#]Joint senior authors, *Corresponding authors

Corresponding authors

Dr Jean Lee: jean.lee@unimelb.edu.au; Dr Ian Monk: ian.monk@unimelb.edu.au

Affiliations:

[1]Department of Microbiology and Immunology, The University of Melbourne at The Doherty Institute for Infection and Immunity, Victoria, Australia

[2]Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The University of Melbourne at The Doherty Institute for Infection and Immunity, Victoria, Australia

[3]Doherty Applied Microbial Genomics, Department of Microbiology and Immunology, The University of Melbourne at The Doherty Institute for Infection and Immunity, Victoria, Australia

[4]Moyne Institute of Preventative Medicine, Department of Microbiology, School of Genetics and Microbiology, Trinity College Dublin, Dublin, Ireland

[5]Infectious Diseases Department, Austin Health, Victoria, Australia

Abstract

Staphylococcus epidermidis is a significant opportunistic pathogen of humans. Molecular studies in this species have been hampered by the presence of restriction-modification (RM) systems that limit introduction of foreign DNA. Here we establish the complete genomes and methylomes for seven clinically significant, genetically diverse *S. epidermidis* isolates and perform the first systematic genomic analyses of the type I RM systems within both *S. epidermidis* and *Staphylococcus aureus*. Our analyses revealed marked differences in the gene arrangement, chromosomal location and movement of type I RM systems between the two species. Unlike *S. aureus*, *S. epidermidis* type I RM systems demonstrate extensive diversity even within a single genetic lineage. This is contrary to current assumptions and has important implications for approaching the genetic manipulation of *S. epidermidis*. Using *Escherichia coli* plasmid artificial modification (PAM) to express *S. epidermidis hsdMS*, we readily overcame restriction barriers in *S. epidermidis*, and achieved transformation efficiencies equivalent to those of modification deficient mutants. With these functional experiments we demonstrate how genomic data can be used to predict both the functionality of type I RM systems and the potential for a strain to be transformation proficient. We outline an efficient approach for the genetic manipulation of *S. epidermidis* from diverse genetic backgrounds, including those that have hitherto been intractable. Additionally, we identified *S. epidermidis* BPH0736, a naturally restriction defective, clinically significant, multidrug-resistant ST2 isolate as an ideal candidate for molecular studies.

Importance

Staphylococcus epidermidis is a major cause of hospital-acquired infections, especially those related to implanted medical devices. Understanding how *S. epidermidis* causes disease and devising ways to combat these infections has been hindered by an inability to genetically manipulate "hospital-adapted" strains that cause clinical disease. Here we provide the first comprehensive analyses of the mechanisms whereby *S. epidermidis* resists the uptake of foreign DNA and demonstrate that these are distinct from those described for *S. aureus*. Until now it had been assumed that these are the same. Using these insights, we demonstrate an efficient approach for the genetic manipulation of *S. epidermidis* to enable the study of clinically relevant isolates for the first time.

Introduction

Staphylococcus epidermidis is a ubiquitous coloniser of human skin (1). Invasive medical procedures, specifically insertion of prosthetic devices on which the bacteria can form a biofilm, enables evasion of both antibiotics and the host immune system and has contributed to its rise as a significant nosocomial pathogen. A leading cause of surgical site and central-line-associated bloodstream infections (2), *S. epidermidis* poses a major economic burden (3). In the hospital environment two multi-locus sequence types (MLSTs), ST2 and ST23, account for most clinical disease (4, 5). Three hospital adapted clones (two ST2, and one ST23) were recently demonstrated to be globally disseminated and have evolved to become nearly untreatable though the acquisition of multiple antibiotic resistance determinants and resistance conferring mutations (5). Knowledge of the molecular genetics, pathogenesis and treatment of *S. epidermidis* has been limited by barriers preventing the genetic manipulation of clinically relevant isolates and the assumption that *S. epidermidis* is similar to *S. aureus*.

Restriction-modification (RM) systems have evolved as a form of bacterial immunity that degrades incoming DNA from foreign donors such as bacteriophage (6). Type I and IV RM systems are a significant barrier to genetic manipulation of staphylococci. Type I RM systems are comprised of three host specificity of DNA (*hsd*) genes that encode (i) a specificity protein (HsdS), (ii) modification proteins (HsdM) and (iii) a restriction endonuclease (HsdR). Together these function as a single protein complex in which HsdS determines the DNA target recognition motif (TRM) in which adenine residues are methylated by HsdM, while HsdR cleaves unmodified and non-self-

84 modified DNA (7, 8). Type IV RM systems consist of a single restriction
85 endonuclease that cleaves DNA with inappropriate cytosine methylation (8).
86 Plasmid artificial modification (PAM) is a method to overcome the barrier
87 imposed by RM systems, where plasmid DNA is passaged through a cytosine
88 methylation deficient *Escherichia coli* host (DC10B) that has been engineered to
89 heterologously express the *hsdMS* system of a staphylococcal strain to be
90 transformed. Plasmid DNA extracted from this *E. coli* host mimics the DNA
91 methylation profile of the target strain, thus enabling introduction of plasmid DNA
92 and subsequent genetic manipulation (9).
93
94 Type I RM systems of staphylococci are best understood in *S. aureus*. The
95 distribution of *hsdS* alleles corresponds with clonal complex (CCs) for the 10
96 dominant *S. aureus* lineages (10). Far less is known about the type I RM systems in *S.*
97 *epidermidis*. A recent study suggested that *S. epidermidis* type I RM systems adhered
98 to lineage specific groupings like *S. aureus*. However, this inference was based on
99 only four new *S. epidermidis* methylomes (11) plus the one methylome that was
100 already characterised; the ST2 reference genome of strain BPH0662 (12).
101
102 Here, we present the first systematic genomic analyses of the type I RM systems in *S.*
103 *aureus* and *S. epidermidis* and demonstrate how this data can be used to predict
104 functionality of type I RM systems and transformational competence of strains. We
105 show that PAM is a highly efficient method to enable genetic manipulation of *S.*
106 *epidermidis*, particularly hospital-adapted isolates that possess multiple functional
107 type I RM systems.
108

Results and Discussion

S. aureus type I RM systems are lineage specific

We began this study by testing the notion that *S. aureus* type I RM systems are lineage specific. We compared 128 publicly available finished *S. aureus* genome sequences (Table S1A) and confirmed the chromosomal location and structure for type I RM systems in *S. aureus* is highly conserved. One hundred and ten genomes had a single *hsdR* and two copies of *hsdMS* with the first in forward orientation located in the alpha pathogenicity island and the second in the opposite orientation, located within the beta pathogenicity island (10, 13) (Figure 1A). The remaining 18 strains possessed *hsdR* and a single copy of forward oriented *hsdMS* in the alpha pathogenicity island. Five of the 128 strains possessed a third type I RM system at a non-mobile chromosomal location downstream of *lacA*. Twenty-three strains carried an additional type I RM system on a mobile genetic element, 22 were on staphylococcal cassette chromosome (SCC) elements (Figures 2 & S1) and one type I RM system on a plasmid (strain HUV05). Single variants of both HsdR (NCBI protein accession WP_000331347.1; n = 127) and HsdM (WP_000028628.1; n = 222) were demonstrated for the type I RM systems situated in non-mobile chromosomal locations, indicating stable vertical inheritance. Interruptions in *hsdR* and *hsdMS* due to horizontal gene transfer were rarely seen in *S. aureus*. A solitary example of *hsdS* truncation due to insertion of a bacteriophage was noted in Sa17_S6. Most changes were due to SNPs leading to amino acid substitutions (N = 35) or nonsense mutations (N = 31) in *hsdS* (Figure 2). See Table 1 for a comparison of *S. aureus* and *S. epidermidis* type I RM systems.

We next established a high-resolution phylogeny using 144,727 core genome single nucleotide polymorphisms (SNPs) for the 128 *S. aureus* genomes covering 40 STs (Figure 2). The occurrence of *hsdMS* genes for each genome was mapped across the phylogeny. A total of 48 HsdS subunits were identified with associated TRMs (Table 2, Table S2A). Although the same *hsdMS* genes were present in genetically distinct lineages, the combination of *hsdMS* genes were conserved within each lineage (Figure 2). For example, the same two HsdMS were present in ST250 and ST254 which are single locus variants of ST8. A notable exception to the lineage specificity were the type I RM systems carried on SCC elements (Figure S1B) that may have been acquired from coagulase negative staphylococci (CoNS). The complete *S. epidermidis* BPH0736 *hsdMSR* (non-disrupted, identical sequence) was observed in four *S. aureus* strains from three different STs (ST5, ST59 and ST338) suggesting gene transfer between the species (Figure S1).

S. epidermidis type I RM systems are carried on mobile genetic elements

Seven complete *S. epidermidis* reference genomes were publicly available at the beginning of this study (Table S1). Of these only BPH0662 (12) and RP62a had characterised type I RM system motifs. However the RP62a methylome was determined independently (11) of the finished genome (14). The methylomes of *S. epidermidis* isolates 1457 (15) and 14.1.R1 (16) confirmed that neither possessed a functional type I RM system, consistent with the absence of *hsdM* genes. To improve understanding of the type I RM systems in *S. epidermidis* we conducted PacBio SMRT sequencing and established complete genomes and adenine methylomes for six additional *S. epidermidis* strains from ST2, ST5, ST59 and ST358 and we re-sequenced RP62a (ST10) (See Table S3 for metadata).

The typical chromosomal arrangement of the type I RM system in *S. epidermidis* is shown for the BPH0736 (ST2) (Figure 1B). Unlike *S. aureus* (Figure 1A), in *S. epidermidis* type I RM systems are arranged as a complete three gene operon in either an *hsdRMS* or *hsdMSR* organisation, unless interrupted (Figure S1). Analyses of the 11 closed *S. epidermidis* genomes containing type I RM systems demonstrated their co-occurrence with cassette chromosome recombinase (*ccr*) genes (with or without the presence of *mec*). The 16 type I RM systems present within these 11 genomes were located a mean distance of 11.5 kb (minimum 2.3 kb, maximum 51.0 kb) from the nearest *ccr* (Figure S1A). Similarly, in the 22 *S. aureus* genomes with a SCC-associated type I RM system, the mean distance between *hsdRMS/MSR* and the nearest *ccr* was 6.9 kb (minimum 1.6 kb, maximum 20.9 kb) (Figure S1B).

Cassette chromosome recombinases typically integrate at *orfX* (corresponding to the last 15 nucleotides of the rRNA large subunit methyltransferase (17)) that is located at 31.6 kb in the *S. epidermidis* chromosome (33.3 kb in *S. aureus*). This is the start of a highly plastic region of the chromosome, in which multiple antibiotic resistance genes, drug transporters and insertion sequence (IS) elements have accumulated (12) (Figure S1). All 11 *S. epidermidis* and 22 *S. aureus* with *ccr*-associated type I RM systems were integrated at *orfX* (Figure S1). For type I RM system variants present in multiple isolates, conservation of genes surrounding the system and *ccr* was observed, consistent with the mobilisation of an entire element (Figure S1). Preserved cassette structure between isolates and both species (Figure S1B) led us to hypothesise that the movement of type I RM systems in *S. epidermidis* is mediated by *ccr*, enabling mobilisation on SCC elements between strains and to other staphylococcal species. Localisation in this region of the genome also

predisposes *S. epidermidis* type I RM systems to disruption, potentially rendering variants restriction deficient. This is seen with interruption of *hsdR* by IS elements in BPH0736 (Figure 1B).

S. epidermidis type I RM systems are not strictly conserved within lineages

To expand the *S. epidermidis* dataset we added short read data for 234 publicly available *S. epidermidis* genomes to the 13 finished genomes (Table S1). Unlike *S. aureus*, variability was noted in both the HsdR and HsdM subunits for *S. epidermidis*. Across the 247 genomes, 183 intact HsdR were identified, with five major variants (<90% amino acid pairwise identity threshold) (Table S2). The two variants of HsdR in strain BPH0662 shared only 22% amino acid identity. Similarly, 178 complete HsdM genes were identified with six major variants (Table 3; Table S2). The amino acid sequences of these six variants were markedly divergent. The two variants of HsdM present in BPH0662 shared only 31% amino acid identity.

A maximum likelihood phylogeny for the 247 *S. epidermidis* genomes, derived from 83,210 core-SNPs, sampled from 72 STs was established and the 31 different *S. epidermidis* HsdS subunits identified were overlaid (Figure 3). Where known, their associated TRMs are shown in Table 3 with NCBI protein accession numbers (Table S2). The distribution of *S. epidermidis* HsdS within the population differed markedly from that observed within *S. aureus*, with no strict concordance to lineage specificity. For example, the HsdS from BPH0723 (BPH0723-S; Table 3) was present in 13 isolates from five STs (ST5, ST21, ST46, ST210, and one unclassified), while BPH0662-S2 was identified in 52 isolates from three STs (ST2, ST23 and ST35). Although a high proportion of ST2 isolates shared the same predicted methylome

(Figure 3), the majority of these were known to be clones of internationally disseminated, multidrug resistant strain BPH0662 (5). However, even within this highly clonal group (n = 36), some predicted methylation variation existed. For example, BPH0662-S2 was absent from two isolates, six isolates (including BPH0662) had a truncation in the 12228-S orphan system while two isolates were missing the 12228-S orphan system completely. Furthermore, within ST2, seven different variants of HsdS were identified in 11 arrangements, including the absence of any type I RM system (Figure 3). Of the 247 *S. epidermidis* genomes analysed, 38% did not contain any *hsdS* alleles and were predicted to be restriction deficient.

Our genome sequencing of RP62a_UoM indicated the presence of a single type I RM system with a G**A**GN$_7$TAC TRM (Table 3). Although this motif was consistent with that previously reported by Costa *et al*., the three additional motifs previously described (lacking apparent associated genes) were not detected by our methylome analysis; the low complexity of the motifs (e.g. GGBNNH) and low frequency of detected methylation (12-29%) (11) suggest these may have been artefacts rather than true motifs. Similarly, the three additional low complexity and frequency motifs reported for VCU036 (11) were probable artefacts. Although the ST type was not specified, VCU036 that shared the same methylome as ST89 isolate NIH051475 was reported as CC89 by Costa *et al.* leading to the conclusion that *S. epidermidis* type I RM systems follows *S. aureus*-like lineage specificity (11). We performed *in silco* MLST by two independent methods and determined VCU036 to belong to ST4. Furthermore, our analysis of the 247 *S. epidermidis* genomes demonstrated VCU036 to be phylogenetically distinct from ST89 (Figure 3).

Overall, our analyses demonstrated that contrary to current assumptions (11) the type I RM systems of *S. epidermidis* do not adhere to the lineage specific distribution observed in *S. aureus*. These differences are attributable to the arrangement of *S. epidermidis* type I RM systems as complete three gene operons that reside within a highly mobile region of the chromosome, the movement of which we hypothesise to be mediated by *ccr*.

Recombinant target recognition domains generate HsdS variants with low conservation of amino acid identity

The structure of a typical type I RM system HsdS allele is shown in Figure S2A, composed of two highly variable target recognition domains (TRDs) flanked and separated by conserved regions (CRs) that collectively determine the TRM to be methylated by HsdM. Recombinant pairings of TRDs result in different variants of HsdS (13, 18). Alignments for the range of *S. aureus* and *S. epidermidis* HsdS identified in this study are shown in Figures S2B and S3 respectively. Within our *S. aureus* and *S. epidermidis* collections, 77 variants of HsdS that shared only 24% pairwise identity were identified. This low conservation poses a potential challenge to the high throughput bioinformatic screening for HsdS variants within genomic datasets. However, using the HsdS from ATCC 12228 as the reference translation with our described method, we were able to detect the partial if not complete presence of all HsdS variants in both species. Of note, 12228-S was the only HsdS variant found within both species that clustered with the majority of *S. aureus* variants. In comparison, RP62a-S only captured 18 of the 31 *S. epidermidis* HsdS variants and fragments of under half the *S. aureus* HsdS variants.

258 **_S. epidermidis_ HsdS variants will only interact as part of a specific complex**

259 The arrangement of some *S. epidermidis* type I RM systems, with the presence of a
260 truncated *hsdS* between complete *hsdR* and *hsdMS* genes, suggested that
261 recombination of component genes occurs (e.g. SepiBPH0662I, SepiRP62aI and
262 SepiBPH0704I; Figure S1). Analyses of the 247 genomes indicated that each variant
263 of *hsdS* in *S. epidermidis* is always associated with a specific *hsdR* and *hsdM* gene,
264 with a conserved gene arrangement (unless interrupted), frequently with the same
265 surrounding genes in association with *ccr* (Figure S1A). These observations support
266 our hypothesis for a role for SCC elements in the mobilisation of *S. epidermidis* type I
267 RM systems.
268
269 The presence of an orphan *hsdS* without a partner *hsdM* gene in *S. epidermidis*
270 introduces additional complexity to the prediction of type I RM system functionality.
271 This was demonstrated by 12228-S, the most prevalent HsdS within the dataset, that
272 is found in 64 *S. epidermidis* (Table 3, Figure 3) and three *S. aureus* (Table 2, Figure
273 2) isolates. All examples of this *hsdS* variant followed a truncated *hsdR*, without an
274 *hsdM* gene. We determined that 12228-S is only expressed if its specific interacting
275 variant of *hsdM* (BPH0662-M1; WP_002504638.1) is also present (see Table S4 for
276 full explanation). In contrast, conservation of a single variant of *hsdM* present twice
277 within the same *S. aureus* genome provides redundancy for the expression of type I
278 RM methylation. This is consistent with previous findings where the product of a
279 single copy of the conserved *S. aureus hsdM* allele could functionally interact with
280 both CC8 HsdS when heterologously expressed in *E. coli* (9).
281

Plasmid artificial modification to overcome the type I RM systems in *S. epidermidis* provides equivalent transformation efficiency as deletion of functional type I systems

To determine the restriction barrier posed by type I RM systems in *S. epidermidis* and assess the efficiency of PAM as a means of bypassing restriction barriers (Figure 4), Δ*hsdS* mutants and *E. coli* hosts for PAM were constructed for *S. epidermidis* isolates BPH0662, RP62a and BPH0736. Two different plasmids (pRAB11 (19) and pIMAY (8)) were used in transformation experiments, as each carried a different number of TRMs recognised by the type I RM systems present in each isolate (Figure 4D). Clinical ST2 isolate, BPH0662-WT, was found to have an intractable restriction barrier unless both functional type I RM systems were overcome through complete bypass with PAM in an *E. coli* host (Ec_Se662I-II), by deletion of both complete *hsdS* genes (BPH0662Δ*hsdSI*Δ*hsdSII*), or a combination of both approaches (plasmid from Ec_Se662I transferred into BPH0662Δ*hsdSII* or plasmid from Ec_Se662II transferred into BPH0662Δ*hsdSI*) (Figure 4A).

Using our protocol, the type I restriction barrier in RP62a-WT was found to be incomplete. Low numbers of transformants (10^1 CFU/ml) were obtained with plasmid DNA isolated from DC10B, indicating that bypassing the type IV restriction barrier alone was sufficient to allow genetic manipulation of this strain (Figure 4B) as previously demonstrated (8). Complete bypass of the single type I RM system in this isolate with *E. coli* host Ec_SeRP62aI significantly improved transformation efficiency to 10^4 CFU/ml, which was equivalent to complete absence of a functional type I RM system as determined with the RP62aΔ*hsdS* mutant (Figure 4B). In contrast, when expressing the RP62a *hsdMS* genes from a plasmid in DC10B, Costa

307 *et al*, were unable to completely bypass the type I RM barrier. This discrepancy was
308 attributed to additional RM systems with low frequency methylation (11). However,
309 our results showed that only one type I RM system is present in RP62a, suggesting
310 that the heterologous expression of type I RM systems on a plasmid in DC10B rather
311 than from a single copy of the genes integrated into the chromosome may be
312 suboptimal. Previously, we found that plasmid-based expression of *hsdMS* was both
313 unstable and cells were unable to tolerate high level expression required for complete
314 methylation of the target DNA (Monk et al 2015).
315
316 Clustered, regularly interspaced, short palindromic repeat (CRISPR) loci confer
317 sequence directed immunity against phages and other foreign DNA, and are another
318 recognised barrier to horizontal gene transfer in *S. epidermidis* (20). Our analysis of
319 the CRISPR spacers for RP62a (Table S3D) did not demonstrate the presence any
320 targets on pSK236 (5.6 kb) used by Costa *et al*., or on pRAB11 (6.4 kb) or pIMAY
321 (5.7 kb) used in this study, that would account for their lower transformation
322 efficiency (10^2 CFU/ml per 5 µg plasmid DNA) (11) compared to our protocol (10^4
323 CFU/ml per 5 µg plasmid DNA for both pRAB11 and pIMAY).
324
325 Isolate BPH0736-WT was predicted to be naturally restriction deficient due to the
326 interruption of *hsdR* by IS elements (Figure 1B) but was demonstrated to retain
327 functional methylation conferred by an intact *hsdMS* system. Due to the complex and
328 infrequently occurring TRM dictated by the single *hsdS* (Table 3), neither pIMAY nor
329 pRAB11 had any BPH0736-S TRMs present. Therefore, pIMAY bearing the
330 Δ736*hsdS* insert (pIMAYΔ736*hsdS*) was used as this contained three TRMs (Figure
331 4D). BPH0736-WT was functionally confirmed to be restriction deficient with the

same transformation efficiency of 10^4 CFU/ml demonstrated for both BPH0736-WT and BPH0736Δ*hsdS* using plasmid isolated from non-specific *E. coli* host DC10B and PAM tailored mutant Ec_Se736I (Figure 4C). Further supporting our bioinformatic predictions, like BPH0736, ATCC 12228 (truncated *hsdR* and no *hsdM*; ST8) and BPH0710 (truncation at amino acid 81of HsdS; ST2) were both predicted to have no functional restriction barrier. Similar to BPH0736, both these strains were transformable in the order of 10^4 CFU/ml with plasmid isolated from DC10B, suggesting that this was the maximum transformation efficiency expected for our protocol. Clinical ST2 strain, BPH0676, was also predicted to have no restriction barrier with complete absence of a type I RM system, however similar to BPH0662 the maximum transformation efficiency achieved was only 10^3 CFU/ml suggesting inherent strain dependent factors other than type I RM systems impacted on transformation of these isolates e.g. cell wall thickness (21).

Although the above data demonstrates PAM to be an efficient method to overcome the type I restriction barrier of *S. epidermidis*, we observed potential instability with the integration of multiple *S. epidermidis hsdMS* of particular TRMs in a DC10B *E. coli* background. With serial passage of Ec_Se662I-II the transformational efficiency of plasmid isolated from this *E. coli* host into BPH0662 declined from 10^3 to 10^1 CFU/ml despite all other experimental parameters remaining the same. This was not observed for any of the *E. coli* PAM mutants expressing single *hsdMS*, including Ec_Se662I and Ec_Se662II, that maintained high-level methylation (89.65 – 99.90%) of motifs within the genome (12) (Table S3). Illumina sequencing of the Ec_Se662I-II genome confirmed integration of both *hsdMS* at the expected chromosomal sites but loss of approximately half the coding sequence of both *hsdS* genes for the majority of

the population sequenced. This instability was hypothesised to be due to the burden of excessive DNA methylation (10,930 sites from heterologous expression of two BPH0662 *S. epidermidis* type I RM systems in addition to 38,592 sites of endogenous *E. coli dam* methylation) that may interfere with normal cellular function, rendering expression toxic in *E. coli*. The same likely accounts for the poor transformation efficiency reported when using PAM for NIH4008 (100-fold lower than that observed for isolates with only a single type I RM system) by Costa *et al*. (11). NIH4008 possesses the same type I RM systems as BPH0662, without the truncation of the orphan *hsdS* (Figure 3). Furthermore, although stable chromosomal integration of three *S. aureus hsdMS* systems in *E. coli* DC10B (IM93B) has been described by Monk *et al*., decreased efficiency of methylation was observed with only 10,135 of 14,602 total TRM sites demonstrating detectable methylation (9).

Collectively, our current and previous (9, 12) data suggests that DC10B *E. coli* is unlikely to consistently maintain the heterologous expression of staphylococcal type I RM systems in the setting of high frequency methylation (>10,000 sites). This limitation should not impact plasmid transformation for mutant creation by allelic exchange, which theoretically requires only a single transformant. However, should high efficiency transformation be sought (e.g. direct transposon mutant library selection), then suitable strains can be predicted using genomic data to identify restriction deficient isolates, such as our newly described reference isolate BPH0736, a clinically significant, ST2 isolate. Clinical metadata, genome characteristics, CRISPR spacers (when present), *in silico* resistome, Vitek 2 antibiogram for clinically relevant antibiotics and common plasmid selection markers for the seven new

reference isolates and BPH0662 are shown in Table S3. Metadata and sequencing accession for mutant isolates is listed in Table S5.

Phage transduction of plasmid is subject to type I restriction

Phage transduction is an alternate method for the genetic manipulation of *S. epidermidis*. Particularly, *S. aureus* ST395 lineage specific Φ187, that shares wall teichoic acid (WTA) receptors with *S. epidermidis* (22, 23). Dependent upon the incidental packaging of plasmid introduced into restriction deficient intermediary host, *S. aureus* PS187Δ*hsdR*Δ*sauPSI*, with Φ187 phage machinery (24), the method can be used to transduce a number of CoNS but is not universally applicable to all *S. epidermidis* isolates (23). The observed ability of ST395 *S. aureus* to exchange DNA with some CoNS led Winstel *et al.* to conclude that overlap may exist between the DNA methylation of ST395 *S. aureus* and CoNS that share the same WTA receptors (22, 23). Phage Φ187 transduction experiments performed using our WT isolates and Δ*hsdS* mutants for BPH0662, RP62a and BPH0736 are shown in Figure 5 examining the transfer of pRAB11. These experiments demonstrate that even if successfully transduced into a *S. epidermidis* isolate, plasmids are still subject to degradation by type I RM systems if they bear a recognised TRM. However, in BPH0736 (absent type I restriction) or mutant strains in which systems have been rendered inactive, transduced plasmid remains viable. The methylome for ST395 *S. aureus* has not been characterised, however the draft genome sequence for PS187 (GCA_000452885.1) indicates that both type I RM systems in this isolate are identical to those in *S. aureus* isolate JS395 (ST1093, belonging to CC395 (25)). We predicted the methylome of the isolate to include G**A**GN$_6$TCG (same as AUS0325-MS2) and another unknown TRM (Figure 2, Table 2). The results of our experiments and analyses of the diversity of *S.*

406 *epidermidis* type I RM systems suggest that successful phage transduction of some *S.*
407 *epidermidis* isolates with Φ187 is more likely related to the absence of a functional
408 system, rather than the presence of a shared methylome with ST395 *S. aureus*. This is
409 further supported by experiments performed by Winstel *et al.*, (23) in which Φ187
410 could only transduce pTX15 (26) into RP62a, but not pKOR1 (27). Based on our
411 characterised RP62a TRM, we determined that pTX15 possesses no RP62a motifs,
412 while both pRAB11 and pKOR1 each bear four motifs explaining why neither
413 plasmid is transducible into RP62a-WT.
414
415 Temperature sensitive plasmid, pIMAY, is frequently used for allelic exchange in
416 staphylococci due to the presence of inducible *secY* antisense counter selection, and
417 the lower likelihood of unintended mutations (such as occur with pKOR1) as
418 integrants are selected at 37^{o}C instead of 43^{o}C (28). However, we found that Φ187
419 was not capable of transducing pIMAY into any of the tested strains, including the
420 Δ*hsdS* mutants and naturally restriction deficient BPH0736. We hypothesised this was
421 due to the low copy number of pIMAY in staphylococci, resulting in low levels of
422 incidental packaging of the plasmid within Φ187, as compared to high copy number
423 plasmid pRAB11. Other limitations to Φ187 transduction include a recommendation
424 to use plasmids <10 kb (24), that should not have impacted on pIMAY (5.7 kb),
425 which is smaller than pRAB11 (6.4 kb). Although a simplified harvesting and
426 infection protocol was used compared to that described by Winstel *et al.* (24), in
427 restriction deficient *S. epidermidis* strain BPH0736 we achieved an efficiency of 10^{4}
428 transductants per ml, equivalent to their anticipated results of 10^{1}-10^{4} (24). Of note,
429 the efficiency of the Δ*hsdS* mutants, BPH0662Δ*hsdSI*Δ*hsdSII* and RP62aΔ*hsdS* were
430 both two-log lower compared to BPH0736 (Figure 5), further supporting the presence

of strain dependent factors beyond the barriers posed by type I RM systems and WTA in these backgrounds.

Conclusions

Our results demonstrate marked differences between the type I RM systems in *S. aureus* and *S. epidermidis* that had hitherto been assumed to be same (11). These differences are predominantly attributable to the arrangement and genome location of the *S. epidermidis* type I system as a complete three-gene operon, we hypothesise to be mobilised by *ccr*. Localisation of the operon in a highly plastic region of the chromosome increases the likelihood of horizontal transfer of these complete systems between *S. epidermidis* as well as to other staphylococci. This results in a lack of lineage specificity and higher probability of spontaneous interruption of component genes. This is in contrast to *S. aureus* where the type I systems are typically arranged as one *hsdR* and two *hsdMS* genes located distant from one another in stable regions of the chromosome. The diversity of *S. epidermidis* type I RM systems that do not strictly adhere to ST/CC groupings indicates that genetic manipulation of *S. epidermidis* requires tailoring to each isolate of interest. Attempting transformation without genomic analysis of the methylome could be successful, as our analyses found that 38% of *S. epidermidis* strains did not possess a type I RM system, and not all systems pose an intractable barrier (e.g. RP62a). However, some isolates such as the internationally disseminated, near pan-drug resistant, clone BPH0662 have complex and absolute type I restriction barriers.

We have demonstrated that PAM using a DC10B *E. coli* host is a simple and effective means to bypass the type I RM barrier in *S. epidermidis*, with a plasmid transfer

efficiency equivalent to a complete absence of type I RM systems. The decreasing cost and ready availability of whole genome sequencing has made the sequencing of isolates planned for mutagenesis and their mutant derivatives commonplace, and a practice that is recommended to ensure the absence of acquired secondary mutations (29). If the genome sequence of an isolate is known, its methylome and ability to be transformed can be predicted as follows: 1. Does the isolate possess an intact type I RM system? If not type I methylation will not be expressed and the isolate should be inherently transformable. 2. Each complete type I RM system within a genome should be functional. For an HsdS protein with known TRMs, presence of the TRMs on a vector will likely prevent transformation. 3. Orphaned, complete *hsdS* genes in the absence of an adjacent *hsdM* may be expressed if their associated *hsdM* allele is present elsewhere in the genome. In view of the above, when designing an *E. coli* PAM host, we recommend including all complete *hsdMS* and any complete orphan *hsdS* genes from the *S. epidermidis* strain to be manipulated, to ensure complete recapitulation of the endogenous type I methylome.

The 247 genomes we analysed are by no means an exhaustive representation of all *S. epidermidis* and additional examples of type I RM systems will undoubtedly be catalogued as further sequencing of this organism is performed. However, this genomic sampling together with our functional data was sufficient to draw the above conclusions. In view of the identified complexities associated with the genetic manipulation of *S. epidermidis*, the reference isolate BPH0736 should prove particularly useful. A clinical ST2 isolate representative of international circulating clones (5), BPH0736 is naturally transformable due to the spontaneous interruption of

480 *hsdR*, rendering it highly amenable to both transformation and phage transduction,
481 making it an ideal strain for future molecular studies.
482
483 **Materials and Methods**
484 **Media and reagents.** Bacterial strains, plasmids and oligonucleotides used in this
485 study are listed in Table S6. *S. epidermidis* were routinely cultured at 37^{o}C in brain
486 heart infusion broth (BHI)(Difco). See Supplementary Methods for detailed
487 description of culture media, antibiotics and enzymes.
488
489 **Genome sequencing and analysis.** Refer to Supplementary Methods.
490
491 **Electroporation.** Early stationary phase cultures (8 h) of *S. epidermidis* grown in 10
492 ml of B Media (BM) were added to 90 ml of fresh, prewarmed BM. Cultures were
493 reincubated to an OD_{600} between 0.8 - 0.9 and chilled in an ice slurry for 10 min.
494 Cells were harvested at 3,900 *xg* for 5 min at 4ºC in a swinging bucket rotor and the
495 cell pellet resuspended in 100 ml of autoclaved, ice-cold water. Centrifugation was
496 repeated, and the pellet resuspended in 50 ml of autoclaved ice-cold water. Cells were
497 centrifuged and successively resuspended in 20 ml, 10 ml then 250 µl of autoclaved
498 ice-cold 10% (weight/volume) glycerol. Equal aliquots (50 µl) were frozen at -80^{o}C.
499 Prior to electroporation, cells were thawed on ice for 5 min, then at room temperature
500 for 5 min. Following centrifugation at 5,000 *xg* for 1 min, cells were resuspended in
501 50 µl of 10% glycerol with 500 mM sucrose (filter sterilised). Pellet paint (Novagen)
502 precipitated plasmid DNA was added to the cells, then cells were transferred into a 1
503 mm electroporation cuvette (Bio-Rad) and pulsed at 21 kV/cm, 100 Ω, 25 µF at room
504 temperature. Routinely, 5 µg of plasmid DNA was used, with concentration

505 determined by fluorometric assay (Qubit 2.0; Life Technologies). Cells were
506 incubated in 1 ml of BHI supplemented with 500 mM sucrose (filter sterilised) at
507 28^{o}C for 2 h prior to plating on BHIA containing chloramphenicol 10 μg/ml.
508
509 **Construction of Ec_Se736I and Ec_SeRP62aI *E. coli* hosts.** *E. coli* mutants
510 expressing the relevant *S. epidermidis* type I RM systems in a DC10B background
511 were created using the primers listed in Table S6 as previously described (8, 9, 12).
512 Detailed methodology in Supplementary Methods.
513
514 **Construction of *S. epidermidis* Δ*hsdS* mutants.** The pIMAY(Δ*hsdS*) vectors were
515 constructed using amplified by overlap extension PCR (30) with the A/B/C/D primer
516 sets specified for each strain in Table S6; cloning into the pIMAY vector backbone;
517 subsequent cloning of the insert into the vector; mutant selection and screening were
518 conducted as previously described (5). Detailed methodology in Supplementary
519 Methods.
520
521 **Harvesting Φ187 + pRAB11/pIMAY lysate from *S. aureus* PS187Δ*hsdR*Δ*sauPSI***
522 Φ187 containing pRAB11/pIMAY was harvested from *S. aureus*
523 PS187Δ*hsdR*Δ*sauPSI* using a protocol adapted from Winstel (24). See Supplementary
524 Methods for detailed methodology.
525
526 **Φ187 + pRAB11/pIMAY transduction of *S. epidermidis*.** A phage transduction
527 protocol was adapted from Foster (31). Detailed methodology in Supplementary
528 Methods.
529

Accession numbers

The datasets supporting the results of this article are available from NCBI under BioProject No. PRJNA532483 (sequencing and closed genome assemblies) and Figshare (*S. aureus* HsdS; *S. epidermidis* HsdS; *S. epidermidis* HsdM; *S. epidermidis* HsdR).

Funding information

This project was supported by the Royal Australasian College of Physicians, Basser Research Entry Scholarship/Australian Government Research Training Program Scholarship to JYHL; National Health and Medical Research Council of Australia (NHMRC) Senior Research Fellowship to TPS (GNT1105525); NHMRC Practitioner Fellowship to BPH (GNT1105905); and NHMRC Project Grant (GNT1066791).

Acknowledgments

We thank Volker Winstel for supply of Φ187 and *S. aureus* strain PS187Δ*hsdR*Δ*sauPSI*. We also acknowledge Bernhard Krismer for providing the assembled DNA sequence of pTX15.

Author contributions

Project conceived by JYHL, TPS, BPH and IRM. Experimental work performed by JYHL and IRM, with PacBio sequencing performed by GPC and SJP. Bioinformatic analysis performed by JYHL with assistance from TS, RG and AGdS. JYHL, TJF, BPH, TPS and IRM drafted the manuscript; all authors reviewed and contributed to the final manuscript.

555 **References**

556 1. Human Microbiome Project Consortium. 2012. Structure, function and
557 diversity of the healthy human microbiome. Nature 486:207–214.
558 2. Sievert DM, Ricks P, Edwards JR, Schneider A, Patel J, Srinivasan A, Kallen
559 A, Limbago B, Fridkin S, National Healthcare Safety Network (NHSN) Team
560 and Participating NHSN Facilities. 2013. Antimicrobial-resistant pathogens
561 associated with healthcare-associated infections: summary of data reported to
562 the National Healthcare Safety Network at the Centers for Disease Control and
563 Prevention, 2009-2010. Infect Control Hosp Epidemiol 34:1–14.
564 3. Otto M, Chatterjee. 2009. *Staphylococcus epidermidis* – the 'accidental'
565 pathogen. Nat Rev Microbiol 7:555-567.
566 4. Thomas JC, Zhang L, Robinson DA. 2014. Differing lifestyles of
567 *Staphylococcus epidermi*dis as revealed through Bayesian clustering of
568 multilocus sequence types. Infect Genet Evol 22:257–264.
569 5. Lee JYH, Monk IR, Gonçalves da Silva A, Seemann T, Chua KYL, Kearns A,
570 Hill R, Woodford N, Bartels MD, Strommenger B, Laurent F, Dodémont M,
571 Deplano A, Patel R, Larsen AR, Korman TM, Stinear TP, Howden BP. 2018.
572 Global spread of three multidrug-resistant lineages of *Staphylococcus*
573 *epidermidis*. Nat Microbiol 3:1175-1185.
574 6. Casadesús J, Low D. 2006. Epigenetic gene regulation in the bacterial world.
575 Microbiol Mol Biol Rev 70:830–856.
576 7. Roberts RJ. 2003. A nomenclature for restriction enzymes, DNA
577 methyltransferases, homing endonucleases and their genes. Nucleic Acids Res
578 31:1805–1812.

579 8. Monk IR, Shah IM, Xu M, Tan M-W, Foster TJ. 2012. Transforming the
580 untransformable: application of direct transformation to manipulate genetically
581 *Staphylococcus aureus and Staphylococcus epidermidis*. mBio 3 e00277-11.
582 9. Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ. 2015. Complete bypass
583 of restriction systems for major *Staphylococcus aureus* lineages. mBio
584 6:e00308–15.
585 10. Waldron DE, Lindsay JA. 2006. Sau1: a novel lineage-specific type I
586 restriction-modification system that blocks horizontal gene transfer into
587 *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. J
588 Bacteriol 188:5578–5585.
589 11. Costa SK, Donegan NP, Corvaglia A-R, Francois P, Cheung AL. 2017.
590 Bypassing the restriction system to improve transformation of *Staphylococcus*
591 *epidermidis.* J Bacteriol 199:e00271–17.
592 12. Lee JYH, Monk IR, Pidot SJ, Singh S, Chua KYL, Seemann T, Stinear TP,
593 Howden BP. 2016. Functional analysis of the first complete genome sequence
594 of a multidrug resistant sequence type 2 *Staphylococcus epidermidis*. Microb
595 Genom 2:e000077.
596 13. Roberts GA, Houston PJ, White JH, Chen K, Stephanou AS, L. P. Cooper,
597 Dryden DTF, Lindsay JA. 2013. Impact of target site distribution for type I
598 restriction enzymes on the evolution of methicillin-resistant *Staphylococcus*
599 *aureus* (MRSA) populations. Nucleic Acids Res 41:7472–7484.
600 14. Gill SR, Fouts DE, Archer GL, Mongodin EF, DeBoy RT, Ravel J, Paulsen IT,
601 Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R,
602 Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee
603 C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson

604 KE, Fraser CM. 2005. Insights on evolution of virulence and resistance from
605 the complete genome analysis of an early methicillin-resistant *Staphylococcus*
606 *aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus*
607 *epidermidis* strain. J Bacteriol 187:2426–2438.
608 15. Galac MR, Stam J, Maybank R, Hinkle M, Mack D, Rohde H, Roth AL, Fey
609 PD. 2017. Complete genome sequence of *Staphylococcus epidermidis* 1457.
610 Genome Announc 5:e00450-17.
611 16. Lassen SB, Lomholt HB, Brüggemann H. 2017. Complete genome sequence of
612 a *Staphylococcus epidermidis* strain with exceptional antimicrobial activity.
613 Genome Announc 5:e00004-17.
614 17. Boundy S, Safo MK, Wang L, Musayev FN, O'Farrell HC, Rife JP, Archer GL.
615 2013. Characterization of the *Staphylococcus aureus* rRNA methyltransferase
616 encoded by *orfX*, the gene containing the Staphylococcal Chromosome
617 Cassette *mec* (SCC*mec*) insertion site. J Biol Chem 288:132–140.
618 18. Cooper LP, Roberts GA, White JH, Luyten YA, Bower EKM, Morgan RD,
619 Roberts RJ, Lindsay JA, Dryden DTF. 2017. DNA target recognition domains
620 in the Type I restriction and modification systems of *Staphylococcus aureus*.
621 Nucleic Acids Res 7:3395-2406.
622 19. Helle L, Kull M, Mayer S, Marincola G, Zelder M-E, Goerke C, Wolz C,
623 Bertram R. 2011. Vectors for improved Tet repressor-dependent gradual gene
624 induction or silencing in *Staphylococcus aureus*. Microbiology 157:3314–
625 3323.
626 20. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal
627 gene transfer in staphylococci by targeting DNA. Science 322:1843–1845.

628 21. Löfblom J, Kronqvist N, Uhlén M, Ståhl S, Wernérus H. 2007. Optimization of
629 electroporation-mediated transformation: *Staphylococcus carnosus* as model
630 organism. J Appl Microbiol 102:736–747.
631 22. Winstel V, Liang C, Sanchez-Carballo P, Steglich M, Munar M, Bröker BM,
632 Penadés JR, Nübel U, Holst O, Dandekar T, Peschel A, Xia G. 2013. Wall
633 teichoic acid structure governs horizontal gene transfer between major bacterial
634 pathogens. Nat Commun 4:2345.
635 23. Winstel V, Kühner P, Krismer B, Peschel A, Rohde H. 2015. Transfer of
636 plasmid DNA to clinical coagulase-negative staphylococcal pathogens by using
637 a unique bacteriophage. Appl Environ Microbiol 81:2481–2488.
638 24. Winstel V, Kühner P, Rohde H, Peschel A. 2016. Genetic engineering of
639 untransformable coagulase-negative staphylococcal pathogens. Nat Protoc
640 11:949–959.
641 25. Larsen J, Andersen PS, Winstel V, Peschel A. 2017. *Staphylococcus aureus*
642 CC395 harbours a novel composite staphylococcal cassette chromosome *mec*
643 element. J Antimicrob Chemother 72:1002-1005.
644 26. Peschel A, Götz F. 1996. Analysis of the *Staphylococcus epidermidis* genes
645 *epiF*, -*E*, and -*G* involved in epidermin immunity. J Bacteriol 178:531–536.
646 27. Bae T, Schneewind O. 2006. Allelic replacement in *Staphylococcus aureus*
647 with inducible counter-selection. Plasmid 55:58–63.
648 28. Monk IR, Foster TJ. 2012. Genetic manipulation of Staphylococci - breaking
649 through the barrier. Front Cell Infect Microbiol 2:49.
650 29. Monk IR, Howden BP, Seemann T, Stinear TP. 2017. Correspondence:
651 Spontaneous secondary mutations confound analysis of the essential two-

652 component system WalKR in *Staphylococcus aureus*. Nat Commun
653 8:ncomms14403.
654 30. Heckman KL, Pease LR. 2007. Gene splicing and mutagenesis by PCR-driven
655 overlap extension. Nat Protoc 2:924–932.
656 31. Foster TJ. 1998. 7.9 Molecular genetic analysis of staphylococcal virulence.
657 Method Microbiol 27:433-454.
658 32. Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K. 2008. Genome
659 sequence of *Staphylococcus aureus* strain Newman and comparative analysis
660 of staphylococcal genomes: polymorphism and evolution of two major
661 pathogenicity islands. J Bacteriol 190:300–310.
662 33. Roberts RJ, Vincze T, Posfai J, Macelis D. 2005. REBASE--restriction
663 enzymes and DNA methyltransferases. Nucleic Acids Res 33:D230–2.
664

Figure 1. Comparison of the structure and chromosomal location of *S. aureus* and *S. epidermidis* type I restriction modification systems. A. *S. aureus* Newman_UoM (29, 32). **B.** *S. epidermidis* BPH0736. For consistency, the chromosome was orientated forwards starting at *dnaA* and native type I RM systems were sequentially numbered, with any imported systems numbered thereafter.

Figure 2. *S. aureus* native type I restriction modification systems are lineage specific. Maximum-likelihood core-SNP based phylogeny of 128 closed *S. aureus* genomes originating from 40 STs, using Newman_UoM as the reference genome. Overlaid are the results of in silico multi-locus sequence type (MLST), clonal cluster (CC), bayesian analysis of population structure (BAPS), presence of CRISPR-Cas systems and type I restriction modification system HsdS variants. Bold red font indicates isolates with PacBio characterised methylomes; Bold blue font indicates isolates with methylomes determined by DNA cleavage with purified enzyme. Boxes around strain names are coloured according ST type. Open circle = amino acid substitutions present in HsdS. *indicates truncated HsdS subunit. Scale bar indicates number of nucleotide substitutions per site (bold) with an approximation of SNP rate (in parentheses).

Figure 3. *S. epidermidis* type I restriction modification systems are not conserved within lineages. Maximum-likelihood, core-SNP based phylogeny for 247 *S. epidermidis* genomes: seven newly closed reference genomes; six existing reference genomes; 156 genomes curated from NCBI sequence read archive (SRA); 75 isolates from Lee *et al*. (5); and the three draft genomes with methylation data (11). BPH0736 was used as the reference genome for analyses. Overlaid are the results of in silico

multi-locus sequence type (MLST), bayesian analysis of population structure (BAPS), presence of CRISPR-Cas systems and type I restriction modification system HsdS variants. Bold red font indicates isolates with characterised methylomes. Isolates were from 70 recognised and two unclassified MLST groups. Boxes around strain names are coloured according ST type; when background colour is same as BAPS group, indicates an ST represented by a single isolate. *Truncated HsdS subunit. Scale bar indicates number of nucleotide substitutions per site (bold) with an approximation of SNP rate (in parentheses).

Figure 4. Plasmid artificial modification to overcome the type I RM systems in *S. epidermidis*. Biological triplicate data for 5 ug of plasmid passaged through DC10B *E. coli* compared to the relevant *E. coli* PAM construct, transformed into *S. epidermidis* wild type (WT) and Δ*hsdS* mutant strains. Error bars represent mean ± standard deviation of three independent experiments. CFU = colony forming units. *No transformants. **A.** Transformation of BPH0622-WT, BPH0662Δ*hsdS1*, BPH0662Δ*hsdS2* and BPH0662Δ*hsdS1*Δ*hsdS2* with plasmid pIMAY or pRAB11 isolated from DC10B and strain specific *E. coli* Ec_Se662I (expressing BPH0662*hsdMS1*), Ec_Se662II (expressing BPH0662*hsdMS2*) and Ec_Se662I-II (expressing both BPH0662*hsdMS1* and BPH0662*hsdMS2*). **B.** Transformation of RP62a-WT and RP62aΔ*hsdS* with plasmid pIMAY or pRAB11 isolated from DC10B and strain specific *E. coli* Ec_SeRP62aI (expressing RP62a*hsdMS*). **C.** Transformation of BPH0736-WT and BPH0736Δ*hsdS* with plasmid pIMAYΔ736*hsdS* isolated from DC10B and strain specific *E. coli* Ec_Se736I (expressing BPH0736*hsdMS*). Note, pIMAYΔ736*hsdS* was used as neither pIMAY

nor pRAB11 possessed any TRM. **D.** Number of *S. epidermidis* strain specific HsdS TRM present on each plasmid.

Figure 5. *S. epidermidis* phage transduction is subject to type I restriction. Biological triplicate data for phage transduction of Φ187-pRAB11 lysate transduced into *S. epidermidis* wild type (WT) and *hsdS* mutant strains. Error bars represent mean ± standard deviation of three independent experiments. *No transformants.

Table 1. Comparison of *S. aureus* and *S. epidermidis* type I restriction modification systems.

Table 2. Diversity of *S. aureus* type I restriction modification system methylation profiles. Isolate HsdS motifs were collated from publications by Monk *et al.*, 2015 (9), Cooper *et al.*, 2017 (18) and the REBASE database (33). HsdS names in bold black font have motifs determined by PacBio sequencing of the isolate after which the representative HsdS was named. HsdS names in bold blue font have motifs determined by DNA cleavage with purified restriction enzyme. Multi-locus sequence types (MLSTs) in which each HsdS are found are listed according to the order they appear in Figure 2 phylogeny (top to bottom); STs within the same clonal complex (CC) are listed within square brackets; STs within parentheses are single locus variants of the ST group they are listed after. trunc = truncated; **A** (red) = methylated adenine residue; **T** = complementary partner to methylated adenine residue. [a] Truncation at amino acid 203. [b] Truncation at amino acid 249. [c] Truncation at amino acid 8. *HUV05_RMS3 is carried on a plasmid, not integrated in the chromosome.

738 Full amino acid translations of all 48 HsdS variants are accessible at Figshare DOI:
739 10.26188/5cb01fc089ab2.
740
741 **Table 3. Diversity of *S. epidermidis* type I restriction modification methylation**
742 **profiles.** Isolate HsdS motifs were collated from methylomes newly characterised in
743 this study and publications by Lee *et al.*, 2016 (12) and Costa *et al.*, 2017 (11). HsdS
744 names in bold black font have motifs determined by PacBio sequencing of the isolate
745 after which the representative HsdS was named. Multilocus sequence types (MLSTs)
746 in which each HsdS are found are listed according to the order they appear in Figure 3
747 phylogeny (clockwise); ST185 is a single locus variant of ST2. trunc = truncated; **A**
748 (red) = methylated adenine residue; **T** = complementary partner to methylated
749 adenine residue. [a] ATCC 12228 type I RM system is non-functional with truncated
750 *hsdR*, complete *hsdS* and no *hsdM*; all 64 isolates possessed the same incomplete type
751 I RM system; Motif identified based on methylome for NIH4008 due to presence of
752 HsdM capable of interacting with 12228 HsdS. [b] 14.1.R1 type I RM system is non-
753 functional with truncated *hsdR*, complete *hsdS* and no *hsdM*. [c] L1M substitution. [d] 1st
754 81 amino acids truncated. [e] S295P substitution. [f] 11 amino acid substitutions (K26E, I
755 56V, E59K, E171K, K174R, K175T, E178A, I193V, D201N, Y386F, V434I). Amino
756 acid translations of all 31 HsdS variants (DOI: 10.26188/5cb01e8f23305) and their
757 interacting HsdM (DOI: 10.26188/5cb019b506f53) are accessible through Figshare.
758
759 **Figure S1. A hypothesised role for cassette chromosome recombinase (*ccr*) in the**
760 **mobilisation of *S. epidermidis* and *S. aureus* type I restriction modification**
761 **systems. A.** Complete *S. epidermidis* genomes with type I RM systems. **B.** *S. aureus*
762 genomes with imported type I RM systems. Genomes are orientated forwards starting

763 at *dnaA*. #NCBI uploaded genome does not start at *dnaA*. *SEI not classifiable by
764 existing MLST scheme.
765
766 **Figure S2. Alignment of *S. aureus* HsdS variants. A.** Structure of an HsdS allele
767 with conserved regions (CRs) flanking two variable regions known as target
768 recognition domains (TRD1 & TRD2). **B.** Each TRD typically specifies three to four
769 defined base pairs including a methylated adenine residue (red **A**; **T** = complementary
770 partner to methylated adenine residue); with a four to seven base pair non-specific
771 spacer (N) between the two defined halves, collectively these TRDs determine the full
772 target recognition motif (TRM) specified by an HsdS variant. HsdS names in bold
773 black font have motifs determined by PacBio sequencing of the isolate after which the
774 representative HsdS was named. HsdS names in bold blue font have motifs
775 determined by DNA cleavage with purified restriction enzyme. Alignments of the
776 identified variants of *S. aureus* HsdS are shown adjacent to their TRMs, each formed
777 by a different TRD pairing. Scale above alignments indicates the position in the
778 consensus alignment with mean pairwise identity at each site graphed (green = 100%
779 identity; khaki = 30-100%; red <30%). Blue (TRD1) and red (TRD2) outlines
780 highlight examples of TRDs that recur within the alignments and the TRM base pairs
781 they define. Yellow boxes highlight alignments of HsdS imported into *S. aureus* on
782 Staphylococcal cassette chromosome elements.
783
784 **Figure S3. Alignment of *S. epidermidis* HsdS variants.** HsdS names in bold black
785 font have motifs determined by PacBio sequencing of the isolate after which the
786 representative HsdS was named. Target recognition motifs (TRMs) (when known)
787 and amino acid alignments are shown adjacent. Scale above alignments indicates the

788 position in the consensus alignment with mean pairwise identity at each site graphed
789 (green = 100% identity; khaki = 30-100%; red <30%). Red outline highlights an
790 example of a target recognition domains that recurs within the alignments and the
791 TRM base pairs they define. Yellow box highlights the 12228-S alignment, this HsdS
792 variant was found to be present in both *S. aureus* and *S. epidermidis* and shared
793 conserved regions with the *S. aureus* HsdS variants located within stable
794 chromosomal islands (Figure S2).
795
796 **Table S1. Metadata associated with pre-existing *S. aureus* and *S. epidermidis***
797 **sequencing. A.** *S. aureus* reference genomes; **B.** *S. epidermidis* reference genomes;
798 **C.** *S. epidermidis* SRA isolate metadata; **D.** Lee *et al*., 2018 isolate metadata; **E.** Costa
799 *et al*., 2017 isolate metadata.
800
801 **Table S2. *S. aureus* and *S. epidermidis* type I restriction modification systems**
802 **subunit protein accession numbers and references. A.** *S. aureus* HsdS; **B.** *S.*
803 *aureus* HsdR and HsdM; **C.** *S. epidermidis* HsdS; **D.** *S. epidermidis* HsdR and HsdM.
804
805 **Table S3. *S. epidermidis* reference isolates metadata. A.** Clinical metadata; **B.**
806 Closed genome statistics; **C.** PacBio methylation statistics; **D.** CRISPR; **E.** Vitek 2
807 susceptibilities; **F.** Resistome; **G.** Susceptibilities to commonly used plasmid selection
808 markers.
809
810 **Table S4. Determining the HsdM variant that interacts with 12228 HsdS.**
811

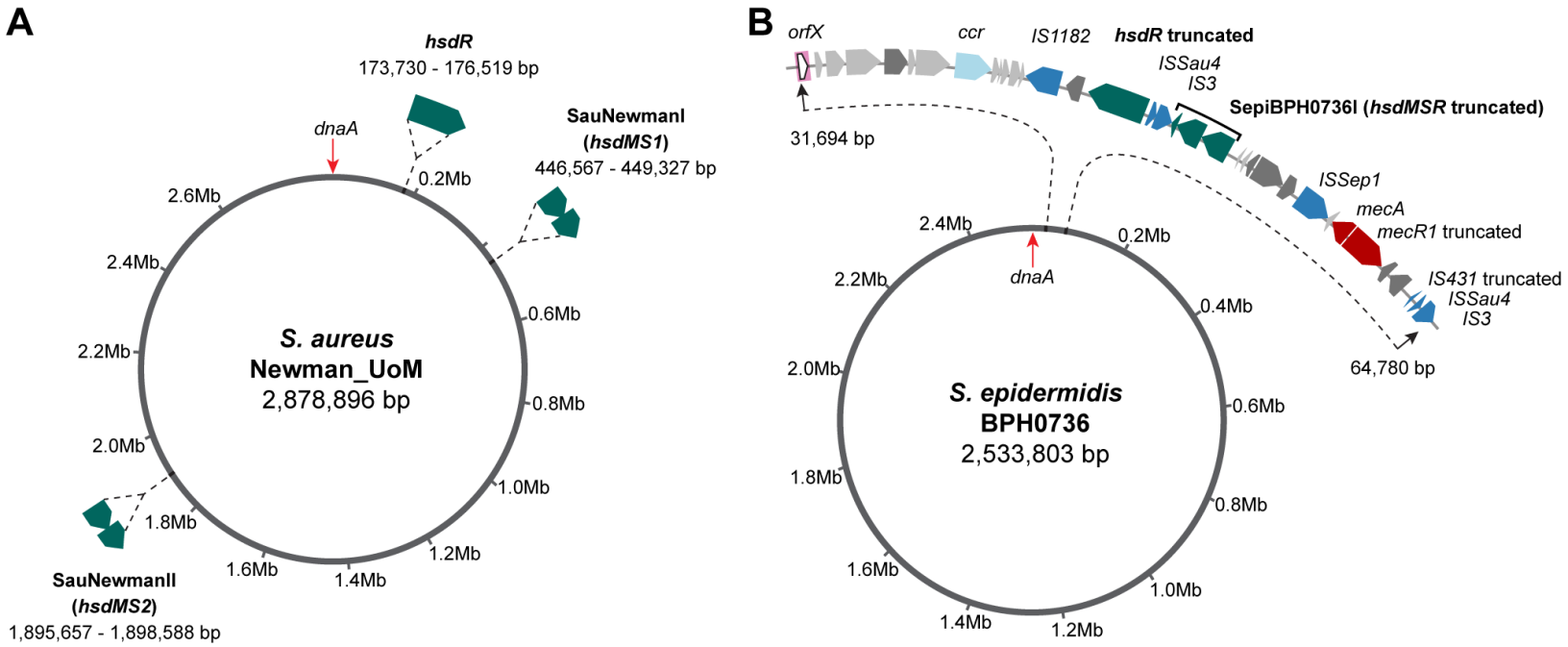
812 **Table S5. *S. epidermidis* Δ*hsdS* and *E. coli* plasmid artificial modification mutant**
813 **sequencing accession data. A.** PacBio sequencing; **B.** PacBio methylation; **C.**
814 Illumina sequencing.

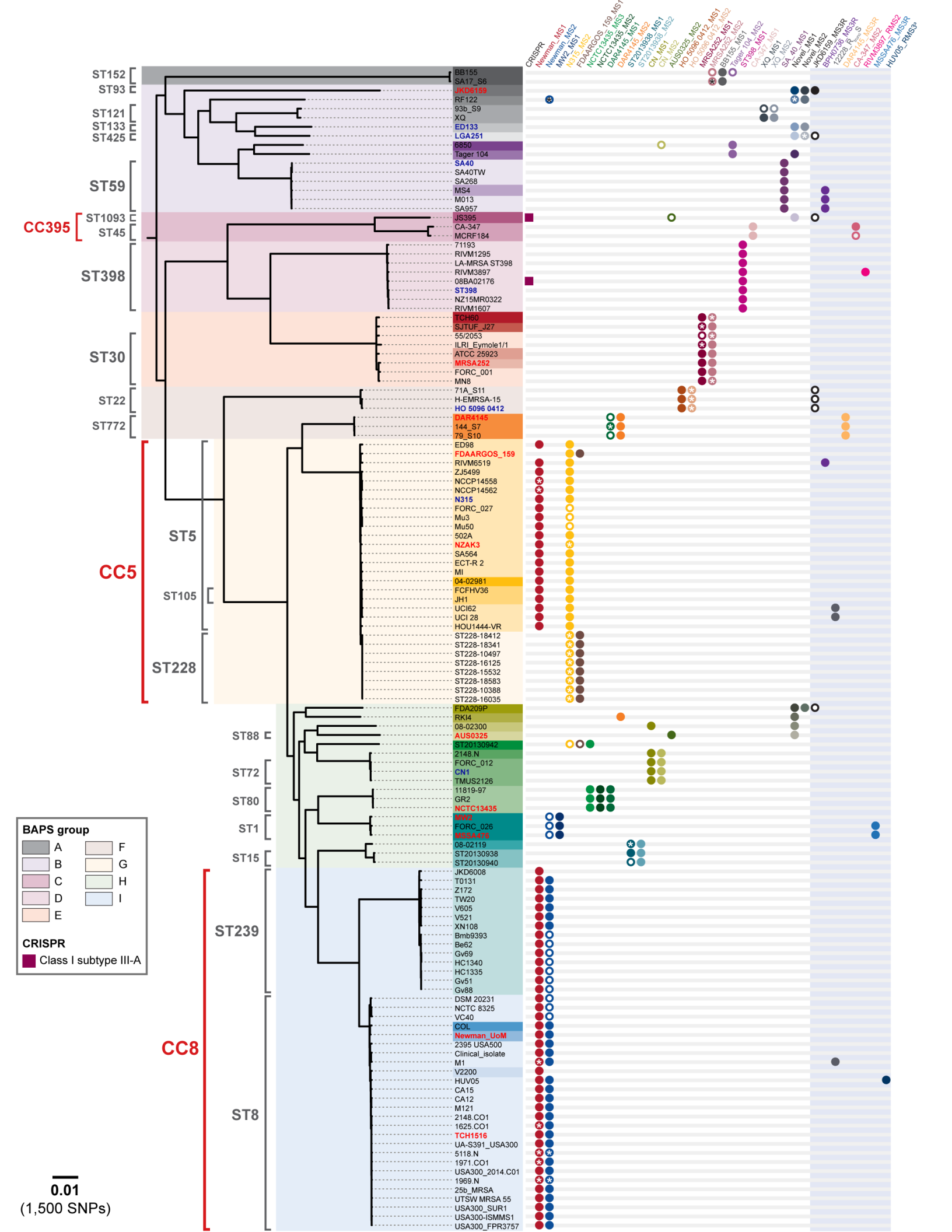
815

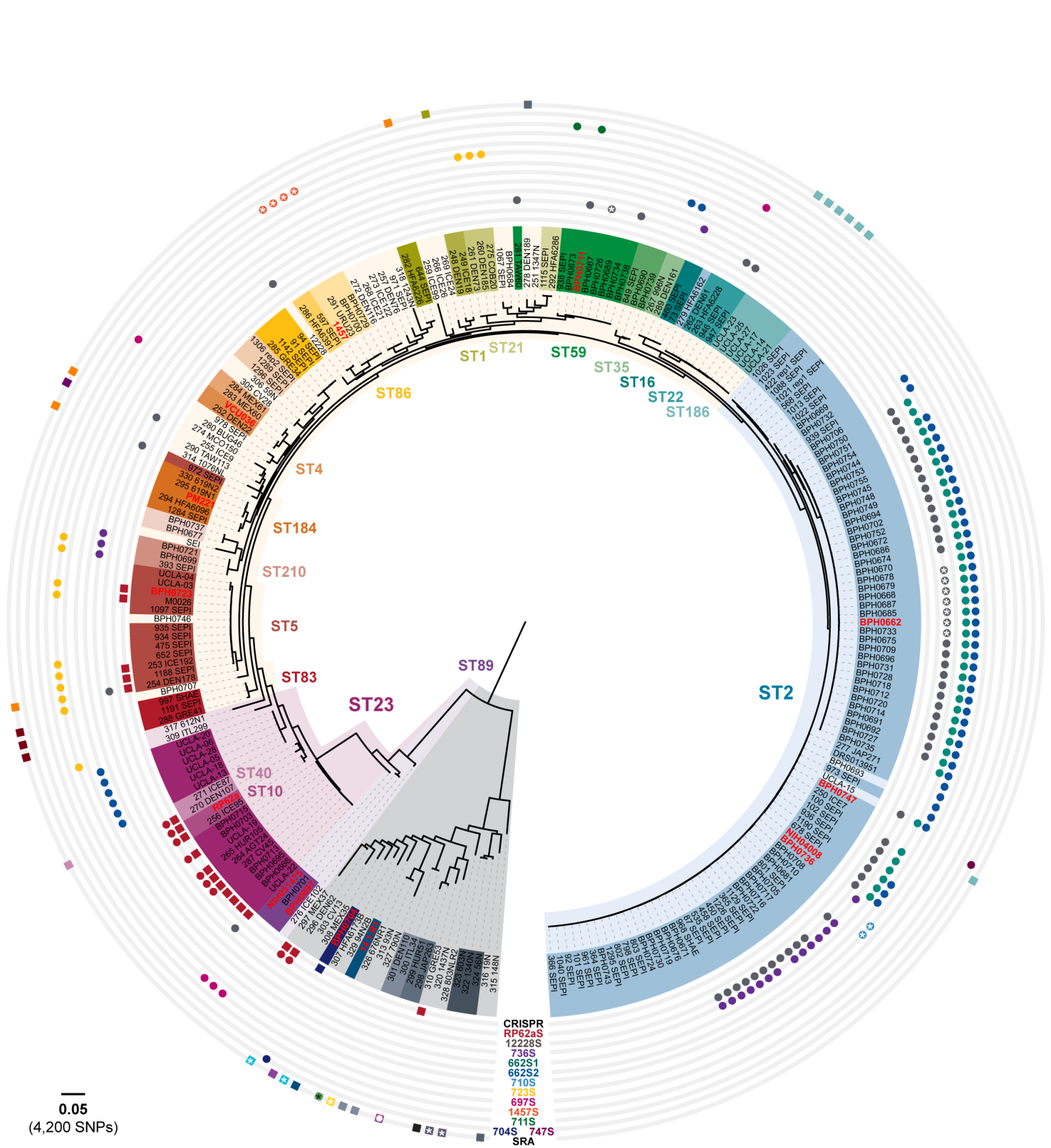
816 **Table S6. Strains, plasmids & oligonucleotides used in this study.**

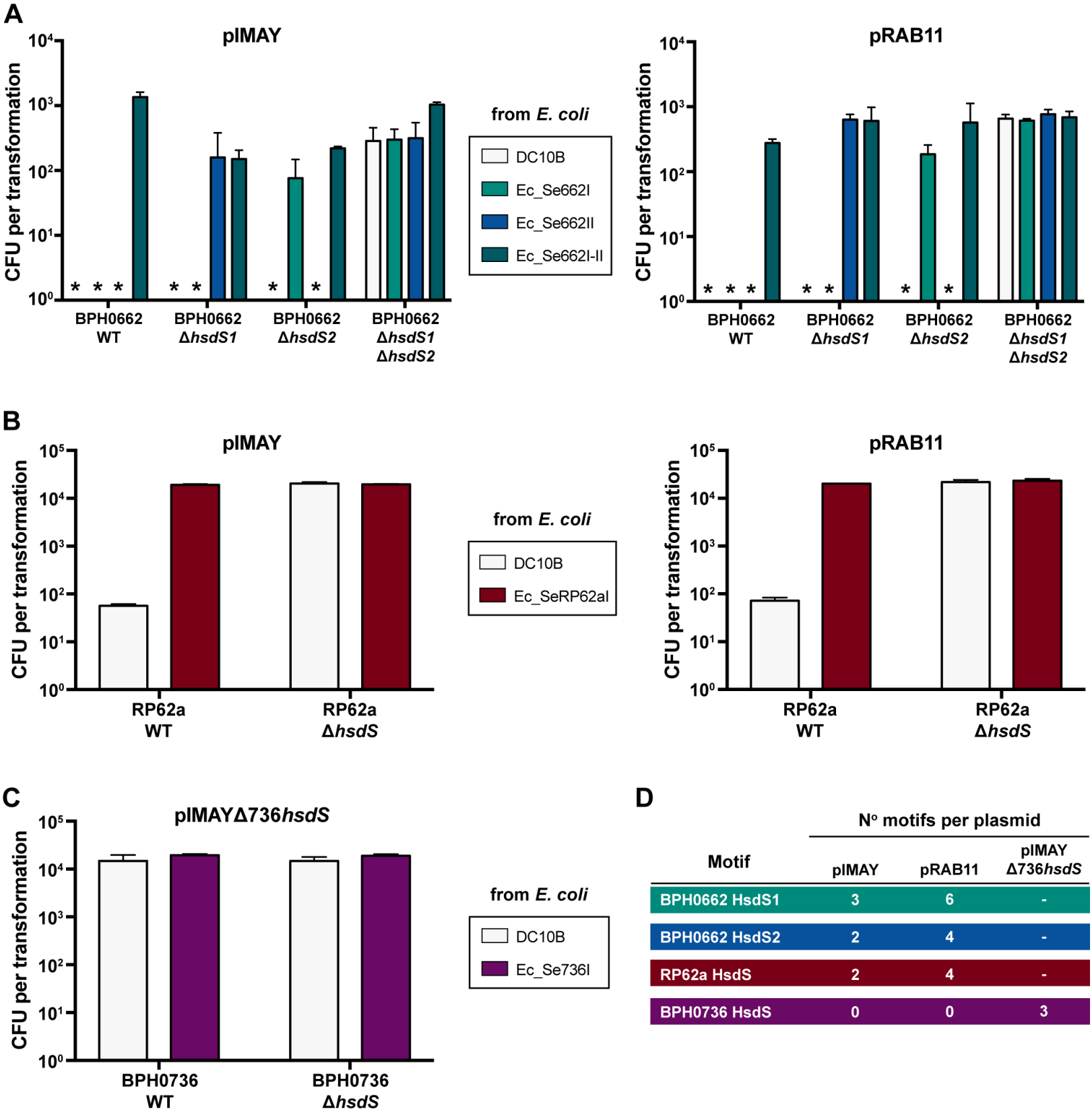
817

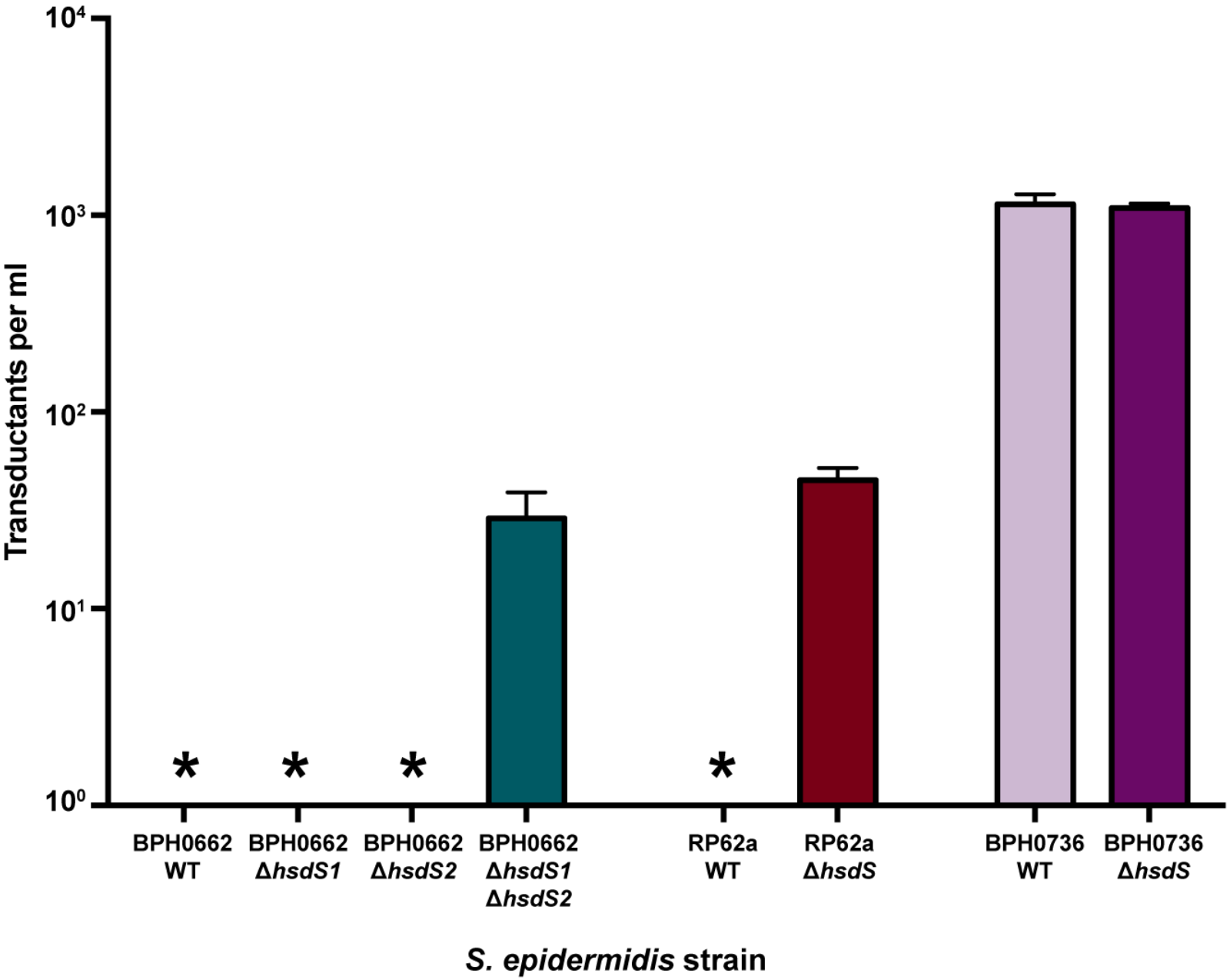
818 **Supplementary Methods.**


A
hsdR
173,730 - 176,519 bp
dnaA
SauNewmanI
(hsdMS1)
446,567 - 449,327 bp
0.2Mb
0.4Mb
0.6Mb
0.8Mb
1.0Mb
1.2Mb
1.4Mb
1.6Mb
1.8Mb
2.0Mb
2.2Mb
2.4Mb
2.6Mb
S. aureus
Newman_UoM
2,878,896 bp
SauNewmanII
(hsdMS2)
1,895,657 - 1,898,588 bp
B
orfX
ccr
IS1182
hsdR truncated
ISSau4
IS3
SepiBPH0736I (hsdMSR truncated)
31,694 bp
ISSep1
mecA
mecR1 truncated
IS431 truncated
ISSau4
IS3
64,780 bp
dnaA
S. epidermidis
BPH0736
2,533,803 bp


CRISPR
Newman_MS1
Newman_MS2
MW2_MS1
N315_MS2
FDAARGOS_159_MS1
NCTC13435_MS3
NCTC13435_MS2
DAR4145_MS1
DAR4145_MS2
ST2013938_MS1
ST2013938_MS2
CN_MS1
CN_MS2
AUS0325_MS2
HO 5096 0412_MS1
HO 5096 0412_MS2
MRSA252_MS1
MRSA252_MS2
BB155_MS1
Tager 104_MS2
ST398_MS1
CA-347_MS1
XQ_MS1
XQ_MS2
SA 40_MS1
Novel_MS1
Novel_MS2
JKD6159_MS3R
BPH0738_MS3R
12228_RpoB_S
DAR4145_MS3R
CA-347_MS2
RIVM3897_RMS2
MSSA476_MS3R
HUV05_RMS3*
ST152
ST93
ST121
ST133
ST425
ST59
CC395
ST1093
ST45
ST398
ST30
ST22
ST772
CC5
ST5
ST105
ST228
ST88
ST72
ST80
ST1
ST15
CC8
ST239
ST8
BB155
SA17_S6
JKD6159
RF122
93b_S9
XQ
ED133
LGA251
6850
Tager 104
SA40
SA40TW
SA268
MS4
M013
SA957
JS395
CA-347
MCRF184
71193
RIVM1295
LA-MRSA ST398
RIVM3897
08BA02176
ST398
NZ15MR0322
RIVM1607
TCH60
SJTUF_J27
55/2053
ILRI_Eymole1/1
ATCC 25923
MRSA252
FORC_001
MN8
71A_S11
H-EMRSA-15
HO 5096 0412
DAR4145
144_S7
79_S10
ED98
FDAARGOS_159
RIVM6519
ZJ5499
NCCP14558
NCCP14562
N315
FORC_027
Mu3
Mu50
502A
NZAK3
SA564
ECT-R 2
MI
04-02981
FCFHV36
JH1
UCI62
UCI 28
HOU1444-VR
ST228-18412
ST228-18341
ST228-10497
ST228-16125
ST228-15532
ST228-18583
ST228-10388
ST228-16035
FDA209P
RKI4
08-02300
AUS0325
ST20130942
2148.N
FORC_012
CN1
TMUS2126
11819-97
GR2
NCTC13435
MW2
FORC_026
MSSA476
08-02119
ST20130938
ST20130940
JKD6008
T0131
Z172
TW20
V605
V521
XN108
Bmb9393
Be62
Gv69
HC1340
HC1335
Gv51
Gv88
DSM 20231
NCTC 8325
VC40
COL
Newman_UoM
2395 USA500
Clinical_isolate
M1
V2200
HUV05
CA15
CA12
M121
2148.CO1
1625.CO1
TCH1516
UA-S391_USA300
5118.N
1971.CO1
USA300_2014.C01
1969.N
25b_MRSA
UTSW MRSA 55
USA300_SUR1
USA300-ISMMS1
USA300_FPR3757
BAPS group
A
B
C
D
E
F
G
H
I
CRISPR
Class I subtype III-A
0.01
(1,500 SNPs)


ST1
ST21
ST59
ST35
ST16
ST22
ST186
ST86
ST4
ST184
ST210
ST5
ST83
ST89
ST23
ST40
ST10
ST2
0.05
(4,200 SNPs)
CRISPR
RP62aS
12228S
736S
662S1
662S2
710S
723S
697S
1457S
711S
704S 747S
SRA
BAPS group
A
B
C
D
E
CRISPR
Class I subtype III-A
Class I subtype II-C
HsdS
RP62a-S
12228-S
BPH0736-S
BPH0622-S1
BPH0622-S2
BPH0710-S
BPH0723-S
BPH0697-S
1457-S
BPH0711-S
BPH0704-S
BPH0747-S


A
pIMAY
pRAB11
CFU per transformation
from E. coli
DC10B
Ec_Se662I
Ec_Se662II
Ec_Se662I-II
BPH0662 WT
BPH0662 ΔhsdS1
BPH0662 ΔhsdS2
BPH0662 ΔhsdS1 ΔhsdS2
B
pIMAY
pRAB11
from E. coli
DC10B
Ec_SeRP62aI
RP62a WT
RP62a ΔhsdS
C
pIMAYΔ736hsdS
from E. coli
DC10B
Ec_Se736I
BPH0736 WT
BPH0736 ΔhsdS
D
N° motifs per plasmid
Motif
pIMAY
pRAB11
pIMAY Δ736hsdS
BPH0662 HsdS1 3 6 -
BPH0662 HsdS2 2 4 -
RP62a HsdS 2 4 -
BPH0736 HsdS 0 0 3


10^4
10^3
10^2
10^1
10^0
Transductants per ml
*
*
*
*
BPH0662 WT
BPH0662 Δ*hsdS1*
BPH0662 Δ*hsdS2*
BPH0662 Δ*hsdS1* Δ*hsdS2*
RP62a WT
RP62a Δ*hsdS*
BPH0736 WT
BPH0736 Δ*hsdS*
S. epidermidis strain

Staphylococcus aureus	***Staphylococcus epidermidis***
Organised as a single *hsdR* gene separated from one, two or three distant *hsdMS* gene pairs	Organised as complete three gene operon (*hsdRMS* or *hsdMSR*)
Conserved, stable chromosomal location for each gene	Close proximity with *ccr* genes integrated at *orfX*, located in highly plastic region of genome
Most have two type I RM systems	Most have a single type I RM system
All have at least one type I RM system	Many (38.1%) have no type I RM system
Up to three functional type I RM systems per isolate	Up to three functional type I RM systems per isolate
99.7% amino acid pairwise identity for all native HsdR	At least five identified variants of HsdR
99.3% amino acid pairwise identity for all native HsdM	At least six identified variants of HsdM
At least 48 different variants of HsdS (eight likely imported from coagulase negative staphylococci)	At least 31 different variants of HsdS
Relative conservation of HsdS present within ST groups	No clear conservation of HsdS according to ST group
Conservation of HsdM provides redundancy, enabling interaction with multiple different HsdS in *S. aureus*	HsdS are only be capable of interacting with their specific paired HsdM, therefore not all orphan *hsdS* will be functional
Complete three gene *hsdRMS/MSR* type I systems carried on SCC elements do not adhere to lineage specificity are likely imported from coagulase negative staphylococci	

	Representative HsdS	Motif	N° isolates	MLST	HsdS NCBI Protein Accession
1.	**Newman_UoM_MS1**	AGG(N)$_5$GA**T**	60	CC5 [5 (225, 105)], CC8 [239, 8 (250, 254, 923)]	WP_000072584.1
2.	**Newman_UoM_MS2**	CCA**Y**(N)$_6$**T**GT	42	151, 1, CC8 [8 (250, 254, 923)]	WP_000072576.1
3.	**MW2_MS1**	CCA**Y**(N)$_5$**TT**AA	3	1	WP_000072566.1
4.	**N315_MS2**	CCA**Y**(N)$_6$G**T**A	30	CC5 [5 (225, 105), 228]	WP_000072627.1
5.	**FDAARGOS_159_MS1**	CCA**Y**(N)$_5$GA**T**	10	CC5 [5, 228], 25	WP_000072629.1
6.	**NCTC13435_MS3**	TCTA(N)$_6$**R**T**T**C	4	25, 72	WP_000072557.1
7.	**NCTC13435_MS2**	GAC(N)$_6$**TTY**G	3	72	WP_000072638.1
8.	**DAR4145_MS1**	CTA(N)$_7$**T**AG	6	772, 72	WP_000072592.1
9.	**DAR4145_MS2**	GAAG(N)$_6$**TTR**G	4	772, 27	WP_000072555.1
10.	ST20130938_MS1	predict ?-R**T**GA	3	582, 15	WP_000066763.1
11.	ST20130938_MS2	predict GGHA-?-**T**TYG	3	582, 15	WP_000072556.1
12.	**CN1_MS1**	GARA(N)$_6$**RT**GT	5	7, unclassified, 72	WP_000072573.1
13.	**CN1_MS2**	GGA(N)$_7$**T**GC	5	50, unclassified, 72	WP_000072558.1
14.	**AUS0325_MS2**	GAG(N)$_6$**T**CG	2	1093, 88	WP_069992008.1
15.	**HO 5096 0412_MS1**	AGG(N)$_6$**T**GAR	3	22	WP_000072565.1
16.	HO 5096 0412_MS2$_{trunc}$	not functional	3	22	WP_000323907.1
17.	**MRSA252_MS1**	GWAG(N)$_5$GA**T**	8	unclassified, 30 (36, 243), 433	WP_000072632.1
18.	**MRSA252_MS2**	GGA(N)$_7$**T**CG	10	unclassified, 30 (36, 243), 433	WP_000072622.1
19.	BB155_MS1	predict AGG-?	2	152	WP_054190421.1
20.	Tager 104_MS2	predict ?-R**T**TC	3	152, 50, 49	WP_000072571.1
21.	**S0385(ST398_MS1)**	ACC(N)$_5$**RT**GA	8	398	WP_000072568.1
22.	**CA-347_MS1**	GWAG(N)$_6$**T**AAA	2	45	WP_000072579.1
23.	XQ_MS1	GGA(N)$_6$CC**T**	2	121	WP_058008355.1
24.	XQ_MS2	GAC(N)$_6$**T**AYG	2	121	WP_058008121.1
25.	**SA40_S**	GGA(N)$_6$**RT**GT	6	59 (338)	WP_000072559.1
26.	**JKD6159_MS1**	CAG(N)$_6$**T**TC	1	93	WP_000072616.1
27.	**JKD6159_MS2**	GGHA(N)$_7$**T**CG	1	93	WP_000072554.1
28.	RF122_MS1	unknown	1	151	WP_000072567.1
29.	RF122_MS2$_{trunc}$	not functional	1	151	WP_070007671.1[a]
30.	**ED133_MS1**	CAG(N)$_5$**RT**GA	1	133	WP_000072617.1
31.	**ED133_MS2**	GGA(N)$_7$**TTR**G	1	133	WP_000072562.1
32.	**LGA251_MS1**	GWAG(N)$_5$**RT**GA	1	425	WP_000072635.1
33.	**LGA251_MS2$_{trunc}$**	not functional	1	425	WP_044122248.1[b]
34.	**Tager 104_S1**	predict GGA-?-R**T**GA	1	49	WP_000072560.1
35.	**JS395_MS1**	predict CAG-?	1	1093	WP_000072352.1[c]
36.	FDA209P_MS1	CCA**Y**(N)$_6$**RT**C	1	464	WP_000072586.1
37.	FDA209P_MS2	predict CCA**Y**-?-**TTY**G	1	464	WP_047210362.1
38.	RKI4_MS1	TCTA(N)$_6$**TT**AA	1	27	WP_000072580.1
39.	08-02300_MS2	predict ?-R**T**TC	1	7	WP_000072588.1
40.	**AUS0325_MS1**	ACC(N)$_5$**RT**GT	1	88	WP_000072613.1
Imported systems					
41.	**JKD6159_MS3R**	GAAG(N)$_5$**T**AC	7	93, 425, 1093, 22, 464	WP_000394004.1
42.	**BPH0736_MS3R**	GAT(N)$_4$CTTA	4	59 (338), 5	WP_000456234.1
43.	**12228_R$_{trunc}$S3**	GAA(N)$_6$CTTA	3	5, 8	WP_001631029.1
44.	**DAR4145_MS3R**	TTAC(N)$_5$TAC	3	772	WP_000394606.1
45.	CA-347_MS2	unknown	2	45	WP_000809140.1
46.	RIVM3897_MS2R	unknown	1	398	WP_060585130.1
47.	MSSA476_MS3R	unknown	2	1	WP_000085808.1
48.	HUV05_RMS3*	unknown	1	8	WP_048520805.1

	Representative HsdS	Motif	Nº isolates	MLST	HsdS NCBI Protein Accession	Interacting HsdM
1.	12228-S[a]	GA**A**$(N)_6$C**T**TA	64	22, 5, 6, 8, 59, 22, 2 (185)	WP_001631029.1	BPH0662-M1
2.	**BPH0662-S2**	C**A**G$(N)_4$A**T**C	52	23, 35, 2	WP_002504701.1	BPH0662-M2
3.	**BPH0662-S1**	**A**TT$(N)_5$C**T**C	43	2	WP_002504637.1	BPH0662-M1
4.	**BPH0736-S**	TA**A**G$(N)_4$A**T**C	18	210, 16, 2	WP_000456234.1	BPH0662-M2
5.	**BPH0723-S**	GA**A**Y$(N)_5$**T**GC	13	32, 46, 83, 5, 210, 21	WP_002469391.1	BPH0662-M2
6.	**RP62a-S**	G**A**G$(N)_7$**T**AC	9	230, 71, 23, 10, 40	WP_002489618.1	BPH0662-M1
7.	**BPH0697-S**	AC**A**$(N)_5$G**T**G	4	89, 4, 22	WP_002505893.1	BPH0697-M1
8.	**BPH0711-S**	GG**A**$(N)_6$**T**AG	2	59	WP_002437522.1	BPH0662-M2
9.	**BPH0704-S**	CYY**A**$(N)_6$CG**T**	1	358	WP_061544233.1	BPH0662-M1
10.	**BPH0747-S**	CN**A**C$(N)_4$R**T**TA	1	2	WP_100481761.1	BPH0662-M2
11.	UCLA-14-S	unknown	7	186, 2	WP_002504281.1	BPH0662-M2
12.	**1457-S_{trunc}**	non-functional	4	86	WP_079118799.1[c]	unknown
13.	972_SEPI-S	unknown	4	5, 6, 20	WP_049387567.1	BPH0662-M2
14.	288_GRE41-S	unknown	3	83	WP_049366366.1	BPH0662-M2
15.	300_ITL34-S	unknown	3	66, 57	WP_021298801.1[d]	BPH0662-M1
16.	308_MEX35-S_{trunc}	non-functional	2	328, 559	WP_080035763.1	unknown
17.	332_1340N-$S1_{trunc}$	non-functional	2	402	novel	BPH0662-M2
18.	332_1340N-S2	unknown	2	402	novel	332_1340N-M
19.	BPH0710-S_{trunc}	non-functional	2	2	WP_064584491.1[e]	BPH0710-M
20.	**14.1.R1-S**[b]	unknown	1	unclassified	WP_102841642.1	unknown
21.	248_DEN19-S	unknown	1	1	novel	BPH0662-M1
22.	271_ICE87-S	unknown	1	40	novel	BPH0662-M1
23.	290_TAW113-S	unknown	1	85	novel	BPH0662-M1
24.	298_JAP60-S1	unknown	1	33	novel	unknown
25.	298_JAP60-S2	unknown	1	33	novel	298_JAP60-M
26.	307_HFA173B-S	unknown	1	37	novel	BPH0662-M2
27.	313_93N-S1	unknown	1	329	WP_104992793.1	BPH0662-M1
28.	313_93N-$S2_{trunc}$	non-functional	1	329	novel	unknown
29.	315_148 N-S	unknown	1	390	WP_107639500.1[f]	BPH0662-M2
30.	327_790N-S	predict ?-A**T**C	1	406	WP_080352270.1	BPH0662-M2
31.	328_803NLR2-S	unknown	1	595	WP_002490392.1	BPH0662-M1