# Sequence analysis reveals genetic exchanges and intraspecific spread of SaPI2, a pathogenicity island involved in menstrual toxic shock

Abhignya Subedi,[1] Carles Ubeda,[1] Rajan P. Adhikari,[1] José R. Penadés[2] and Richard P. Novick[1]

Correspondence
Richard P. Novick
novick@saturn.med.nyu.edu

[1]Skirball Institute and Departments of Microbiology and Medicine, New York University Medical Center, New York, NY 10016, USA

[2]Departamento de Química, Bioquímica y Biología Molecular, CEU-Universidad Cardenal, 19 Herrera, 46113 Moncada, Valencia, Spain

SaPIs are a family of homologous phage-related pathogenicity islands in staphylococci that carry superantigen and other virulence genes, and are responsible for a wide variety of superantigen-related diseases. SaPIs are induced to excise and replicate by particular staphylococcal phages and are encapsidated in infectious, small-headed, phage-like particles, which are transmitted at very high frequency among staphylococcal strains and species. SaPI2 is a prototypical member of this family that was identified in a typical menstrual toxic shock syndrome (TSS) strain of *Staphylococcus aureus*, the so-called Harrisburg strain, and found to be mobilizable by typing phage 80. Most menstrual TSS strains belong to a highly uniform *agr* group III clone of electrophoretic type (ET) 41, and this study was undertaken to determine whether such strains typically carry SaPI2, and whether it has spread beyond the ET41 clone. We report here the complete sequence of SaPI2, describe its relation to other known SaPIs, and show that it, or a very similar element, is carried by most ET41 strains but that it has disseminated to other strains that have also been implicated in TSS. We show additionally, that SaPIs are widespread among the staphylococci and that most TSS strains carry two or more, including SaPI2.

## INTRODUCTION

Shortly after the first description of staphylococcal toxic shock syndrome (TSS) (Todd *et al.*, 1978) and the identification of TSS toxin-1 (TSST-1) as the causative agent (Bergdoll & Schlievert, 1984), it was observed that only 10–20 % of natural staphylococcal isolates produced the toxin (Altemeier *et al.*, 1982). Those that did not produce it lacked the gene *tst*, whereas producers possessed not only the gene but also some 15 kb of additional DNA (Kreiswirth *et al.*, 1989). Chromosomal mapping studies revealed that the TSST-1-encoding element in RN4282 and in a second menstrual TSS prototype, the so-called Harrisburg strain (RN3984 in our strain series), were at different chromosomal locations – the former linked to the *tyr* locus and the latter to the *trp* locus. Since the *tst* genes

Abbreviations: ERP, excision–replication–packaging; ET, electrophoretic type; PG, putative gene; SaPI, *Staphylococcus aureus* pathogenicity island; SEB, enterotoxin B; TSS (TSST), toxic shock syndrome (toxin).

The GenBank/EMBL/DDBJ accession number for the nucleotide sequence of SaPI2 is EF010993.

Additional experimental procedures, tables and figures are available with the online version of this paper.

at both locations were identical, we assumed that the element was a transposon, and designated it Tn*557* (Chu *et al.*, 1988). Southern blotting with a clone containing the RN4282 *tst* and flanking sequences revealed that the two *tst*-carrying elements were not identical and also differed from the corresponding elements in other clinical strains (Lindsay *et al.*, 1997), suggesting that the element was probably not a transposon after all, but rather that each was a member of a family of similar elements. Historically, we concentrated on the *tst* element in RN4282, since that was the strain that had been used for cloning of the gene (Kreiswirth *et al.*, 1984). We sequenced this *tst* element and found that it was about 15 kb in length, encoded two superantigen toxins in addition to TSST-1, as well as integrase and terminase small subunit homologues, and it was flanked by a 17 nt direct repeat. We found that the element is induced to excise and replicate by a temperate phage, 80α, is packaged into small-headed phage-like particles and is transmitted at extremely high frequencies (Lindsay *et al.*, 1998; Ruzin *et al.*, 2001). On the basis of these results, we proposed that the *tst* element is a mobile pathogenicity island and designated the RN4282 prototype superantigen pathogenicity island 1 (SaPI1) and that in

RN3984, SaPI2 (Lindsay et al., 1998). A third member of the SaPI family was identified in a bovine strain and designated SaPIbov (Fitzgerald et al., 2001) (now known as SaPIbov1), and one or more SaPIs have been identified in all but one of the nine sequenced Staphylococcus aureus genomes (Novick & Subedi, 2007). Around the time that SaPIs were first identified, it was shown by Musser et al. (1990) that the vast majority of menstrual TSS strains belong to a single clone, of electrophoretic type (ET) 41, and it became clear with the discovery of the agr groups (Ji et al., 1997) that these strains, including RN3984, were all agr group III, whereas RN4282 was agr group I, was not a member of this clone and was not, in retrospect, a typical menstrual TSS strain. RN3984, harbouring SaPI2, therefore, is the menstrual TSS prototype and therefore merits full characterization.

In this paper, we report the complete sequence of SaPI2 from RN3984, demonstrate its induction and transfer by certain staphylococcal phages, and show that it, or very similar elements, are present in about 70 % of menstrual TSS strains, including some of the strains characterized by Musser et al. (1990). SaPI1 and SaPIbov1 have recently been characterized in sufficient detail to provide a general understanding of the molecular genetics of the SaPIs as a family (Novick & Subedi, 2007). The features of SaPI2 are described in relation to our current understanding of these two.

## METHODS

**Bacterial strains and plasmids.** Bacterial strains used in this study are listed in supplementary Table S1 (available with the online version of this paper). RN4220 is a mutant of RN450 selected for the ability to accept Escherichia coli DNA. RN10615 is a derivative of RN4220, lysogenic for phage $\phi$53. RN4282 and RN3984 are naturally occurring TSS strains and are the original sources of SaPI1 and SaPI2, respectively. RN8652 is a derivative of RN3984 with tetM inserted in tst in SaPI2. RN10619 is a derivative of RN10615, obtained by transduction of SaPI2 tst::tetM from RN8652 with phage 80. S6C is a naturally occurring enterotoxin B (SEB) producer (Jones & Khan, 1986). In addition to the strains listed in Table S1 (available with the online version of this paper) is a set of 43 clinical TSS strains, kindly provided by Drs Schlievert and Musser during the 1980s. These are listed in supplementary Table S2 (available with the online version of this paper). Although these strains are all menstrual TSS isolates, they have not hitherto been characterized for the presence or expression of tst. Some of the strains have been electrophoretically typed, others have not. Additional methods are described in an online supplement (available with the online version of this paper).

## RESULTS AND DISCUSSION

### The SaPI excision–replication–packaging (ERP) cycle in staphylococci

The interaction between temperate phages and SaPIs is quite specific – only a very few (staphylococcal) phages can induce any given SaPI. A key feature of the SaPI ERP cycle is the appearance, late in the lytic cycle, of an SaPI-specific band that migrates ahead of the bulk (sheared chromosomal and phage) DNA in a standard agarose gel (Lindsay et al., 1998). It is suggested that this band represents monomeric SaPI DNA released from mature SaPI particles, as it is not seen in mutants defective in the terminase small subunit (Ubeda et al., 2007a).

### Phage-specific induction of SaPI2

In an earlier study, we observed with the native SaPI2 strain, RN3984, that typing phage 80 could induce the SaPI2 ERP cycle, whereas 80α could not (Ruzin et al., 2001), though the two phages were thought to be closely related. To determine whether this difference in specificity was a host-specific or phage-specific effect, we transduced SaPI2-tst::tetM into RN4220, and attempted to test for induction of the ERP cycle by each of the two phages. Phage 80 did not grow on RN4220 and therefore could not be tested for ERP induction in this strain; 80α, however, which grows very well on RN4220, could induce the ERP cycle for SaPI2-tst::tetM as well as for SaPI1-tst::tetM in this strain (not shown). Both SaPIs caused a ~100-fold reduction in the 80α titre on RN4220 and were transduced at the typical high frequencies of 0.024 for SaPI1 and 0.14 for SaPI2. We hypothesize that the failure of 80α to induce SaPI2 in RN3984 resulted from the necessity to 'adapt' 80α to RN3984. 80α grown on our standard strain, RN450, from which RN4220 was derived, had a plating efficiency of about $10^{-4}$ on RN3984 (~$10^6$ vs ~$10^{10}$ p.f.u. ml$^{-1}$ on RN450), but the phage resulting from propagation on this strain had a plating efficiency of about 1 (~$10^{10}$ p.f.u. ml$^{-1}$ on both strains) – which we assumed was owing to a restriction-modification-based adaptation. In retrospect, it is likely that 80α had, additionally, recombined with an endogenous prophage in RN3984, losing the ability to induce SaPI2 in the process. Consistent with this is the equivalent plating efficiencies of the 'adapted' and native phages; since SaPI induction generally causes a 100-fold reduction in phage titre (Ruzin et al., 2001), if the adapted phage could induce SaPI2, its titre should have been ~100 fold lower. We note that endogenous prophages harboured by RN3984 and at least seven other menstrual TSS strains cannot induce the ERP cycle of their endogenous SaPIs (Ruzin et al., 2001), whereas phage 80 could induce it and there are several examples of endogenous prophages that can induce a resident SaPI (see below and Table 1). Of the known SaPIs, at least seven can be induced either by endogenous or superinfecting phages on the basis of the appearance of an SaPI-specific gel band in whole-cell lysates during phage growth. The genetic basis of phage specificity for SaPI induction, however, remains to be determined.

### SaPI2 sequence

As the above results enabled us to use 80α for SaPI2 induction in RN4220, which contains no endogenous prophage, we isolated an 80α lysogen of RN4220,

**Table 1.** The SaPI family

| Element (strain*) | Size (kb) | Comment | Inducing phages | *att* site core | *att*$_S$ site | Accessory genes | Orient-ation | References |
|---|---|---|---|---|---|---|---|---|
| SaPI4 (MRSA252) | 15.1 | | Endogenous prophage | AAAGAAGAACAATAATAT | 8' | None | + | Holden *et al.* (2004); A. Subedi & R. P. Novick, unpublished |
| SaPI1028 (NY940) | 15.6 | Originally phage PT1028 | Endogenous prophage | AAAGAAGAACAATAATAT | 8' | None | + | Kwan *et al.* (2005); A. Subedi & R. P. Novick, unpublished |
| SaPIbov1 (RF122 *v*Sa2) | 15.8 | Excised sponta-neously | φ11 φ1477, 80α | TAATTATTCCCACTCAAT | 9' | *tst, sel, sek* | + | Fitzgerald *et al.* (2001); Ubeda *et al.* (2003) |
| SaPIbov2 (V329) | 27 | Excised sponta-neously | 80α | TAATTATTCCCACTCGAT | 9' | *bap* | + | Ubeda *et al.* (2003) |
| SaPIm4 (mu50 : SaGIm *v*Sa3 **Type I**) | 14.4 | | Endogenous prophage | TCCCGCCGTCTCCAT | 18' | *fhuD* | + | Baba *et al.* (2002); A. Subedi & R. P. Novick, unpublished |
| SaPImw2 (mw2 : *v*Sa3 **Type II**) | 14.4 | | Endogenous prophage | TCCCGCCGTCTCCAT | 18' | *ear, se2l, sec4* | + | Baba *et al.* (2002); A. Subedi & R. P. Novick, unpublished |
| ShPI2 (*S. haemolyticus, v*Sh2) | 16.6 | | ND | TCCCGCCGTCTCCAT | 48'† | None | − | Takeuchi *et al.* (2005) |
| SaPI1 (RN4282 *v*Sa1) | 15.2 | | 80α, φ13 | TTATTTAGCAGGAATAA | 19' | *ear, tst, sek, seq* | + | Lindsay *et al.* (1998) |
| SaPI3 (COL *v*Sa1) | 15.6 | | 29 (?) | TTATTTAGCAGGAATAA | 19' | *ear, seb, sel, sek* | + | Novick *et al.* (2001) |
| SaPI5 (USA300) | 14.0 | | ND | TTATTTAGCAGGAATAA | 19' | *ear, sek, seq* | + | Diep *et al.* (2006) |
| SaPIn1 (n315) SaPIm1 (mu50) (*v*Sa4 **Type I**) | 15 | Identical | ND | GTTTTACCATCATTCC-CGGCAT | 44' | *tst, sel, sec3* | − | Kuroda *et al.* (2001) |
| SaPI2 (RN3984) | 14.7 | | 80, 80α | ATTTTACATCATTCCT-GGCAT | 44' | *tst, eta* | − | Ruzin *et al.* (2001); A. Subedi & R. P. Novick, unpublished |
| SaPI122 (RF122) | 17.9 | | Endogenous prophage | GTTTTACATCATTCCT-GGCAT | 44' | *mdr* | − | GenBank NC_007622 |
| SaPIfusB | 16.7 | | ND | TTTTTACATCATTCCT-GGCAT GTTTTACATCATACCT-GGCAT | 44' | *fusB* | − | GenBank AM292600; O'Neil *et al.* (2007) |
| SaPI6Δ (8325, COL, USA300, MSSA476, mw2 : *v*Sa4 **Type II**) | 3.14 | Identical | ND | GTTTTACCATCATTCC-CGGCAT GTTTTACATCATTCCT-GGCAT | 44' | None | − | Baba *et al.* (2002); R. P. Novick, unpub-lished |
| SsPI15305 (*S. saprophyticus* 15305, *v*Ss15305) | 17.0 | | ND | CGAGGGGACTAATAA-GT CGAGGGGATTAATAA-GT | 47' | *aad, fosB* | − | Kuroda *et al.* (2005) |

ND, No data.

*Strain designation is from Baba *et al.* (2002).

†ShPI2 is located 180° away from the other SaPIs having the same *att* core sequence, owing to the major chromosomal inversion that has been documented in the *S. haemolyticus* genome.

introduced SaPI2-*tst* :: *tetM*, prepared an 80α lysate, concentrated the phage by centrifugation and then phenol-extracted the phage DNA. This preparation, treated with proteinase K and then ethanol-precipitated, consisted of >80 % SaPI and <20 % phage DNA (not shown). It was used without further purification as template for the sequencing of SaPI2-*tst* :: *tetM*. Sequencing was accomp-lished by primer walking in both directions from the *tst*

gene; the final sequence was 14 755 bp long. We note that one short sequence in SaPI2 matched a sequence in 80α, which we have determined independently (A. Matthews & others, unpublished). This is the ~50 nt binding site for the regulatory proteins RinA and RinB (Ye & Lee, 1993), which is conserved among all of the SaPIs, including SaPI2; since it is flanked by SaPI sequences rather than 80α sequences, it is clearly endogenous to SaPI2 rather than representing phage DNA contamination. Thus, the presence of 80α DNA in the material used for sequencing was apparently irrelevant.

## SapI2 genome organization

To place SaPI2 in the context of similar elements found among staphylococcal strains, we will review briefly the present status of SaPI studies. Sequences are now available for 17 full-length SaPIs plus a 3.1 kb SaPI remnant, SaPI6Δ, present in five of the sequenced genomes. In general, the SaPIs form a coherent group with a conserved overall genome organization (Novick & Subedi, 2007). The conserved core genome, transcriptional organization, and mobility mechanism indicate common ancestry and serve to separate the SaPIs from all other mobile or potentially mobile elements among the staphylococci. We have recently identified SaPI-like elements in the lactococci (R. P. Novick, unpublished) and suggest that such elements are an important feature of the mobile genome in Gram-positive bacteria.

The SaPI genome has two components – genes involved in the ERP cycle and those contributing to virulence or other phenotypic functions. The latter, accessory genes, are located in any of three regions – A1, at the extreme left end, A2, immediately to the right of the integrase gene (*int*), and A3, at the extreme right end (Table S3, available

with the online version of the journal). Several of the known SaPIs do not contain identifiable accessory genes. In Fig. 1 the overall organization of the SaPI genome (generic) is shown along with the specifics of SaPI2 genome organization, individual ORFs, designated SP2-PG (SaPI2 Putative Gene) in the text and indicated as numbers above. Where functions are known, 'PG' is replaced by 'G'. Gene abbreviations are indicated below the SaPI2 map. The suggested SaPI nomenclature involves the designation of each of the ERP genes belonging to a particular SaPI with a capital letter referring to that element. Thus SaPI1 genes are 'A', SaPIbov1 genes are 'B', SaPI2 genes are 'C', etc. We have not included the 'C' in Fig. 1 to save space.

In most SaPIs, including SaPI2, *int* is at the extreme left end. Some other SaPIs have 1 or 2 genes to the left of *int*. In SaPI2, *int* is immediately followed by a conserved regulatory module defining two major divergent transcription systems, headed by regulatory genes, *stl* (SaPI transcription leftward) and *str* (SaPI transcription rightward). Other SaPIs may have as many as four PGs between *int* and the regulatory module, including known toxin genes. The first set of genes in the rightward transcription system of SaPI2 is involved in replication and consists of two genes of known function, *pri* (primase) and *rep* (replication initiation), preceded by four smaller PGs of unknown function. Other SaPIs have two to five of the latter. Immediately 3′ to *rep* is the (SaPI2-specific) replication origin (*ori*), present at the same location in all SaPIs (Ubeda *et al.*, 2007a). Following *ori* are three SaPI2 PGs of unknown function followed by a second major operon, operon I, that contains six genes involved in forming the SaPI-specific capsids (Ubeda *et al.*, 2007a). Several of the SaPIs lack one or more of these genes; in one case, that of SaPIbov2, small capsids are not produced and the SaPI DNA is packaged in full-sized phage heads
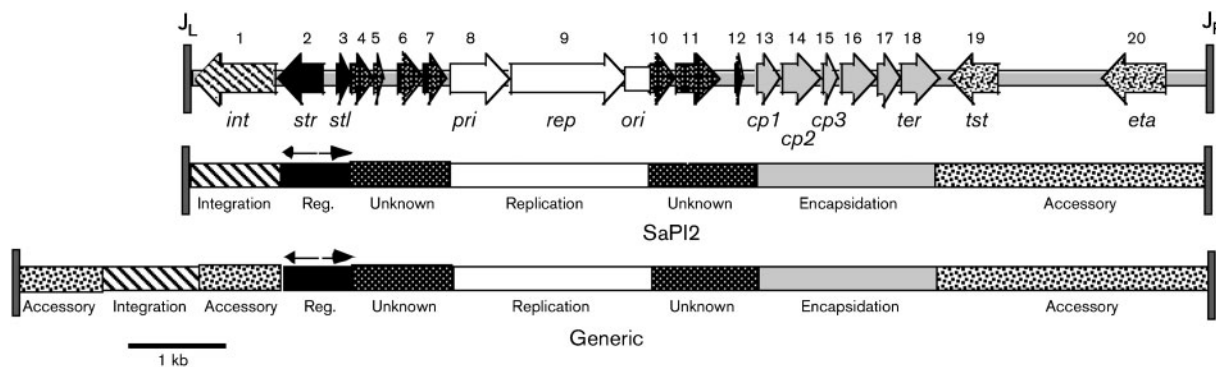


**Fig. 1.** Schematic SaPI2 map. A summary diagram of the generic SaPI genome with different hatching patterns representing different functions is presented at the bottom of the figure. The *int* region is on the left, flanked by accessory gene regions. These are followed by the regulatory module, a set of PGs of unknown function, the replication operon, another set of PGs of unknown function, the encapsidation operon and, finally, a third accessory gene region. These modules are always in the same order, with variable regions of accessory genes or putative genes of unknown functions interspersed between them. The middle diagram is of the SaPI2 genome, showing the corresponding regions, and a representation of the SaPI PGs, using the same patterning, is shown at the top.

(Maiques *et al.*, 2007). The last gene in operon I, *ter*, is the terminase small subunit homologue, absolutely required for the encapsidation of SaPI DNA (Ubeda *et al.*, 2007a). To the right of *ter*, in the variable A3 region, are *tst* and a gene, *eta*, encoding an exfoliatin A homologue (see below).

SaPI2 thus contains homologues of eight genes known to be involved in the ERP cycle, 10 PGs of unknown function and two accessory genes located in the A3 region. Its overall G + C content, like that of the others, is 31 %, suggesting that the SaPIs originated in staphylococci or closely related organisms. Table 2 shows the similarities between each of the SaPI2 PGs and the corresponding PGs in all of the other known SaPIs. The other two best characterized SaPIs, SaPI1 (15 252 bp) and SaPIbov1 (15 891 bp) are slightly larger than SaPI2. SaPI1 has similar ERP genes but has four accessory genes located in two regions, with only *tst* being shared. SaPIbov1 also has four accessory genes, all in the A3 region; again, only *tst* is shared.

## Accessory genes

The toxin most frequently encoded by SaPIs, including SaPI2, is TSST-1. In fact, SaPIs are exclusively responsible for TSS; one of the questions addressed here is whether *agr* group III strains, which are specifically implicated in menstrual TSS (Musser *et al.*, 1990), share SaPIs similar or identical to SaPI2.

SaPI2, unlike any of the other SaPIs for which sequences are available, contains a homologue of the phage-carried *Staphylococcus hyicus* exfoliatin A gene (*eta*) (Sato *et al.*, 1999). SaPI2 *eta* is ~90 % identical to the *S. hyicus* gene, but both are completely unrelated to the classical *S. aureus eta* and *etb* (whether the *S. aureus* and *S. hyicus* toxins are serologically related is unknown). Remarkably, the predicted SaPI2 and *S. hyicus eta* products are ~90 % identical to the predicted product of a hitherto undescribed *eta* homologue present in all of the sequenced *S. aureus* genomes, always located at about 1.1–1.2 Mb, which is not near any known mobile element in any of the strains. This striking similarity suggests that the obvious source of SaPI2 *eta* is the chromosomal homologue – indeed, there are matching 48 bp regions of SaPI2 and the staphylococcal chromosome 5′ to the *eta* coding sequence. It is therefore likely that toxin genes, such as *eta* and other accessory genes, have been acquired by SaPIs by recombination.

## *att* site sequences and chromosomal locations

The SaPI1 and SaPIbov1 *att* sites were identified as short direct repeats at the junctions between conserved chromosomal flanking sequences and inserted DNA, by comparison of a strain that contained an SaPI with one that did not (Lindsay *et al.*, 1998; Fitzgerald *et al.*, 2001). The SaPI2 *att* site core was identified in the SaPI2 sequence. Since SaPI DNA in transducing particles is circularly permuted (Ruzin *et al.*, 2001), identification of

**Table 2.** Relationships between SaPI2 genes and those of other SaPIs

SaPI2 PG similarities were determined with BLASTX. Shading: light grey, >50 % similarity to corresponding gene in SaPI2; dark grey, 20–50 %; black, <20 %; white, no corresponding gene. Abbreviations: HP, hypothetical protein; *int*, integrase; *stl*, SaPI transcription leftward; *str*, SaPI transcription rightward; *pri*, primase homologue; *rep*, replication initiator, helicase homologue; *cp1* and *cp2*, determinants of capsid size; *ter′*, terminase small subunit homologue; *ter*, terminase small subunit.

| Gene | | SaPI m1/n1 | SaPI 122 | SaPI fusB | SaPI bov1 | SaPI bov2 | SaPI 3 | SaPI 1 | SaPI 5 | SaPI mw2 | SaPI m4 | SaPI 4 | SaPI 1028 | ShPi 2 | SsPI 15305 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1 | *int* | 96 | 96 | 88 | 34 | 36 | 26 | 26 | 26 | 26 | 25 | 30 | 30 | | 37 |
| G2 | *stl* | 100 | | | | | | | | | | | | | |
| G3 | *str* | 100 | | 32 | | 43 | | | | 29 | | | | | |
| PG4 | HP | 100 | 96 | 80 | | | | | | | 27 | | | | |
| PG5 | HP | 84 | 73 | | 76 | 29 | 80 | 71 | 86 | 63 | 71 | 86 | 78 | | |
| PG6 | HP | | | | 97 | | | | | 98 | 96 | | | | |
| PG7 | HP | 43 | 94 | 38 | 96 | 17 | 89 | 86 | 35 | 41 | 92 | 46 | 48 | | 39 |
| G8 | *pri* | 95 | 94 | 37 | 93 | 62 | 78 | 81 | | | 83 | 95 | 99 | 53 | |
| G9 | *rep* | 92 | 92 | 83 | 96 | 82 | | | | | | 93 | 100 | | |
| PG10 | HP | 96 | 96 | | 98 | 78 | 29 | 29 | 22 | | 29 | 96 | 100 | | 29 |
| PG11 | HP | 94 | 94 | 70 | 94 | 74 | 93 | 96 | | | 90 | 93 | 94 | | 64 |
| PG12 | HP | 100 | 92 | | 77 | | 95 | 100 | | | 100 | | 95 | | |
| PG13 | HP | 99 | 98 | 49 | 97 | 52 | 99 | 99 | | | 99 | 96 | 98 | | 51 |
| G14 | *cp1* | 99 | 93 | 44 | 93 | 50 | 100 | 99 | | | 99 | | 43 | | |
| G15 | *cp2* | 98 | 95 | | 95 | | 97 | 100 | | | 98 | | | | |
| G16 | *cp3* | 100 | 94 | 25 | 94 | 31 | 100 | 98 | 33 | 33 | 100 | 95 | 95 | 39 | 42 |
| PG17 | *ter′* | 97 | 95 | | 91 | | 98 | 99 | | | 99 | 94 | 96 | 30 | 32 |
| G18 | *ter* | 98 | 95 | 58 | 96 | 73 | 96 | 98 | 24 | 25 | 98 | 97 | 98 | 52 | 35 |
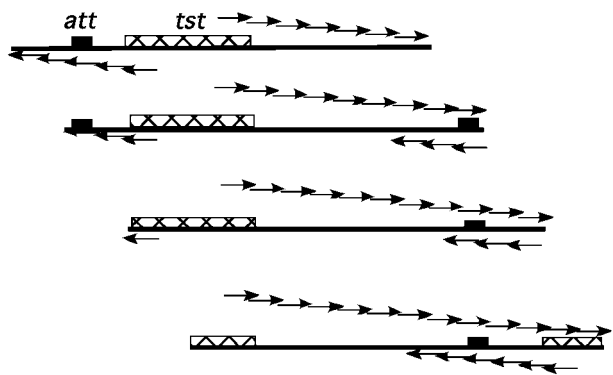
**Fig. 2.** Mapping the SaPI2 $att_S$ core. Several linear permutations of SaPI2 are shown, with *tst* cross-hatched and $att_S$ in black. The primers (arrows) used for sequencing by walking outward from *tst* in both directions eventually come to matching sequences, and the site where these meet represents the $att_S$ core.

the SaPI2 core *att* site was straightforward. As shown in Fig. 2, when sequences obtained with primers facing in one direction match those obtained with primers facing in the opposite direction, the core *att* site then represents a short overlap between the two end regions. As listed in Table 1, this is a 21 nt sequence, 5′-ATTTTACATCATTCCTGG-CAT-3′, that matches the core *att* site for several other SaPIs (see below). A search of the sequenced *S. aureus* genomes revealed that the SaPI2 *att* site core was present in all nine, in some cases with a single nucleotide mismatch. In eight of the genomes, there were two copies, flanking an

SaPI or SaPI-like element. In MRSA252 there was a single copy with no adjacent SaPI-like sequences.

In all cases thus far analysed, one or two copies of any SaPI1 *att* site are present at corresponding locations; where there are two sites, an SaPI or an SaPI-related sequence is present between them. The chromosomal SaPI *att* sites are referred to as $att_C$, and the matching SaPI site, $att_S$. As with classical temperate phages, the SaPI *att* sites contain a conserved core, of 15–22 nt, plus essential flanking sequences that differ between SaPI and chromosome. The SaPI1-specific *att* sites, $att_C$ and $att_S$, are each less than 400 nt in length, on the basis of the functionality of cloned segments (R. P. Novick & A. Ruzin, unpublished). No SaPI *att* site, however, has yet been mapped in detail.

## Integrase specificity and *att* site location

The complex of integrase and insertion site is highly evolved and would not be readily amenable to evolutionary modification, whereas other genetic segments can readily be acquired, exchanged, rearranged or lost. Therefore, chromosomally located mobile genetic elements, such as SaPIs, SCCmecs and prophages, that encode and utilize site-specific integrases, are always located at an integrase-determined site and are divided into subclasses according to integrase/*att* site specificity. The known SaPIs occupy six different sites in the staphylococcal genome, five in *S. aureus* and one in *S. saprophyticus*, and therefore represent six different subclasses. Given the diversity of the sequenced strains, it is likely these represent all or most of the SaPI sites in *S. aureus*. An unrooted tree for the 15 core SaPI genomes is presented in Fig. 3(a), with the *att*/*int*
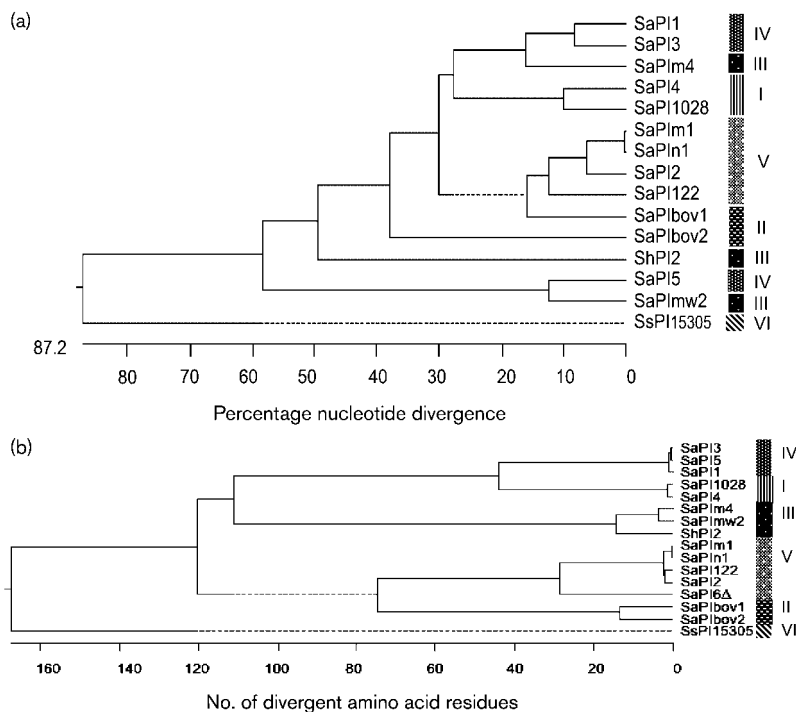


**Fig. 3.** SaPI homologies. SaPI sequences, aside from that of SaPI2, were obtained from GenBank. (a) Nearest neighbour tree for 15 SaPI core genomes. Alignment was performed on the raw SaPI nucleotide sequences, with the known accessory genes removed and a nearest neighbour unrooted tree calculated with VectorNti. The scale represents the distance in percentage divergent nucleotides. *att* site locations are indicated on the right, coded according to Figs 3(b) and 4. (b) Nearest neighbour tree for SaPI integrases. Amino acid alignment was performed and the tree was calculated with VectorNti. The scale represents the distance in number of divergent amino acyl residues. The different $att_C$ site groups are indicated on the right.

types indicated. As can be seen, SaPIs within any *att/int* subclass are generally much more closely related to others in the same subclass than to those in different subclasses. Some important exceptions, however, can be discerned by comparing Figs 3(a) and 3(b), which shows an unrooted nearest-neighbour tree for the SaPI integrases; these almost certainly represent recombinational exchanges as described below. The *att* sites are indicated in Fig. 4 as minutes on a 60 min clock representing the composite staphylococcal genome, and the *int/att* subclasses are numbered I–VI in clockwise order. The SaPI2 *att* site is occupied by SaPI122, the nearly identical elements, SaPIn1 and SaPIm1, and the most recently identified member of the family, SaRIfusB (GenBank accession no. AM292600) (see Table 1 and Fig. 4). This site is also occupied, in five of the nine sequenced strains, by a defective SaPI remnant known as SaPI6Δ, so that the SaPI2 site is occupied in all of the sequenced strains, and it is occupied in 40 of the 43 TSS strains described below. The other *att* sites depicted in Fig. 4 are much less frequently occupied, though they are present in all genomes sequenced thus far. SaPIs in the right replichore are all oriented in the clockwise direction (+), those in the left replichore are all oriented counter-clockwise (−), as is the case with prophages. Note that

ShPI2 is located 180° away from SaPImw2 and SaPIm4, which have the same *att* site core sequence, and is oriented in the opposite direction. This is because a very large segment of the *Staphylococcus haemolyticus* chromosome is inverted with respect to the replication origin (Takeuchi *et al.*, 2005). We note that *S. saprophyticus* contains the *att* site found at 44' in all of the sequenced *S. aureus* genomes, but this site, at the 3' end of *groEL*, is located at 21' in *S. saprophyticus*, whereas the SsPI15305 *att$_C$* site, at 47' (Fig. 4), is also conserved in *S. aureus* but is at about 17', suggesting that there is a major chromosomal rearrangement in this species, which may be the same as that described by Takeuchi *et al.* (2005) in *S. haemolyticus*, as compared to *S. aureus*. This would suggest that the rearrangement was ancestral to the divergence that led to these two species.

Remarkably, as noted above, SaPI1 and SaPI2 were mapped some years ago by three-factor crosses using transformation and transduction, and were reported to be linked to the *tyr* and *trp* loci, respectively (Chu *et al.*, 1988), and, for SaPI2, confirmed more recently by transduction linkage analysis (N. Kurepina & R.P. Novick, unpublished). However, neither of the SaPI1 or SaPI2 *att* sites, identified
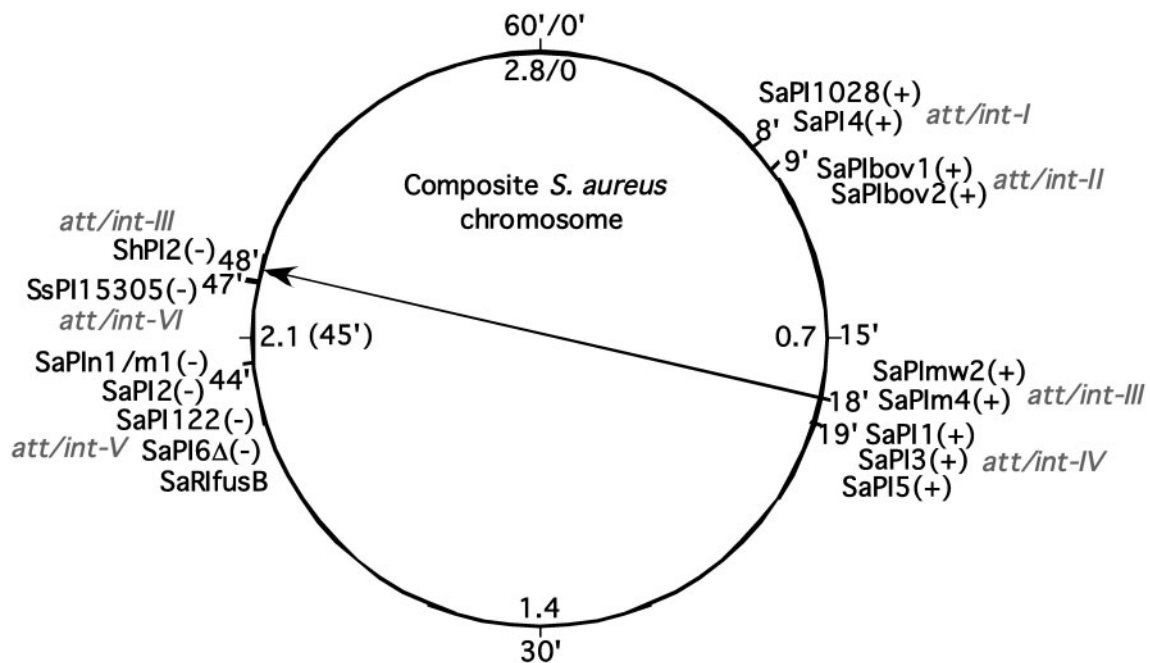


**Fig. 4.** Chromosomal locations of the SaPI *att$_C$* sites. A composite diagram of the staphylococcal chromosome, compiled from published genome sequences, is shown with average locations given in minutes (outer scale) and in Mb (inner scale). The SaPI *att$_C$* sites are numbered clockwise, with SaPIs occupying each site grouped. Note the arrow crossing the diagram; this is meant to indicate that ShPI2, from *S . haemolyticus*, is at the same site, as revealed by its flanking sequences, as the two SaPIs at 18'; the reason for the different location on the circular map is that the *S. haemolyticus* genome, in comparison to that of *S. aureus*, contains a major rearrangement in which about half of its chromosome has been inverted with respect to the other half (Takeuchi *et al.*, 2005). There also appears to be a major translocation in the *S. saprophyticus* genome in comparison to that of *S. aureus*, since the *att$_C$* site for SsPI15305, plus its flanking sequences, at 47', are conserved and are located at approximately 17' in the *S. aureus* genomes.

by genome sequencing and indicated in Fig. 4, are anywhere near either of these two loci. The SaPI1 *att* site has been shown directly to be the same as that of SaPI3, which has been located in the COL chromosome by genome sequencing and confirmed by junctional PCRs. SaPI2 has been convincingly mapped at the indicated site by junctional PCRs. Thus, we have no explanation for this discrepancy.

## The SaPI2 ERP cycle

For SaPI1, and probably SaPI2, the phage-encoded *xis* function is required in addition to integrase. SaPI integrase is required for replication and encapsidation as well as for excision and, not surprisingly, must be expressed in the recipient for establishment of an incoming SaPI (P. Barry & R. P. Novick, unpublished). Interestingly, SaPIbov1 and SaPIbov2, alone among the SaPIs in the sequenced genomes, encode a putative *xis* function (Maiques *et al.*, 2007) and can excise spontaneously in the absence of vegetative phage (Ubeda *et al.*, 2003). SaPI excision is presumably by the Campbell mechanism (Campbell, 1969), and is evidently followed by phage-like replication. The known helper phages, with one exception, are generalized transducing phages and therefore use the *pac*-dependent headful packaging mechanism. Replication is initiated at a unique origin, characterized by a series of hexa- to octanucleotide repeats, by the Rep protein, which is SaPI-specific (Ubeda *et al.*, 2007b), and therefore corresponds to the replicon-specific initiator proteins of other types of replicons. The SaPI1 and SaPI2 *ori*'s are compared in Fig. 5 (Ubeda *et al.*, 2007b) and we have shown independently that the cloned *pri-rep* genes of SaPI1, SaPIbov1 and SaPI2, can drive replication of a plasmid containing the cognate *ori*, but not of any non-cognate one (Ubeda *et al.*, 2007b).
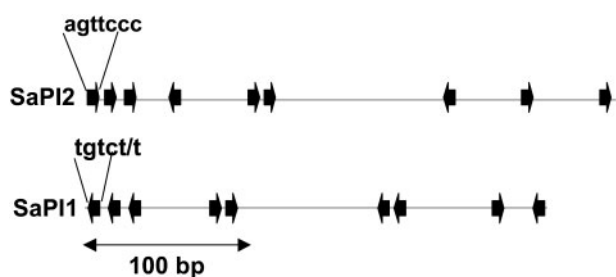


**Fig. 5.** Comparative arrangement of origin repeats for SaPI2 and SaPI1. The SaPI replication origins are located 3′ to the *rep* gene and consist of a series of short oligonucleotide repeats that differ among the different SaPIs. Arrows represent individual repeats of the indicated sequence within the overall 250–300 bp origin region (Ubeda *et al.*, 2007b).

## Role of SaPI2 and its close relatives in menstrual TSS

On the basis of the above description, it is proposed that SaPI individuality may be determined by *int/att* and *rep/ori* specificities. Given that menstrual TSS strains form a well defined clone (Musser *et al.*, 1990), we wished to know whether this clonality is reflected in SaPI individuality: do such strains typically carry SaPI2 or a close relative, and has SaPI2 spread to TSS strains that do not belong to the ET41 group? For this part of the study we selected 50 clinical TSS isolates from a set of about 180 obtained from TSS cases and kindly provided by various colleagues over the years (Table S2). Seven of these proved unsuitable, either because they were not *S. aureus* or because they were duplicates. To gain a preliminary appreciation of the diversity of TSS strains, we included a small set of ET41 strains (11 in all) to provide a biotype baseline, and a larger number of strains that were either untyped or were not ET41. Strains were examined for certain phenotypic characters which are important features of their biotypes and facilitate comparisons: *agr* type represents a broad evolutionary category and is important in this context because all of the ET41 strains are *agr* group III; pigmentation is important because it is roughly correlated with *tst* expression (unpublished observations); haemolytic activity is listed because it is a well known indication of virulence and of *agr* functionality. Several key SaPI features were determined so as to characterize the resident SaPIs in relation to the SaPI2 prototype : : *att* site occupation, integrase homology, *rep-ori* specificity, and for the expression of *tst*. These tests involved Southern blotting of chromosomal DNAs with probes specific for the SaPI2 replication origin, the five different *S. aureus* integrase classes, plus PCR with primers specific for the SaPI2 junctions. PCR was also used to confirm the presence of *tst* and Northern blotting was used to test for *tst* expression. The results of these tests are summarized in online supplementary Table S5 (available with the online version of this paper), and the supporting Southern blotting patterns are shown in supplementary Figs S1–S7 (available with the online version of the journal). It is notable that integrases of four of the five types scored were present, but that no strain had a type II integrase, which has been found thus far only in SaPIs from bovine strains (Ubeda *et al.*, 2003).

## SaPI analysis

We were surprised to find that although all 11 of the ET41 strains expressed *tst*, 11 of the others showed neither a *tst* band by Southern blot hybridization with a *tst* probe nor a TSST-1 band on a Western blot. Two of these strains, with very weak PCR reactions for *tst*, were positive for *seb*, the enterotoxin B gene, by using PCR. As *tst* and *seb* do not appear ever to be co-expressed, it seemed important to follow up the possibility of joint expression of these two superantigens in the context of SaPI content of TSS strains. Since PCR for *seb* was positive, as was a Western blot for

SEB, it is suggested that the initial weak *tst* PCRs were due to cross-reactions with *seb* and that none of the strains analysed does in fact contain or express both *tst* and *seb*. The remaining nine strains, which do not detectably express *tst* or *seb*, may have caused TSS by means of other superantigens (these were not characterized in this study). The majority of menstrual TSS strains are *agr* group III (Ji *et al.*, 1997), including all the ET41 strains studied, 29 in all. Some of the strains, however, belonged to other *agr* groups: of the 43 strains under study, 10 were *agr*-I, three were *agr*-II and one was *agr*-IV.

Any strain that matched SaPI2 by integrase type, junctional sequences and replication origin specificity was considered to contain an SaPI2-like element. Of the 43 strains, 26 were positive for SaPI2 by these criteria, including the 11 ET41 strains. Of the 26 SaPI2 strains, 20, including all 11 of the ET41 strains, were positive for type I integrase as well as for type V, suggesting that they contain a second SaPI, belonging to *att/int* subclass I, possibly similar to either SaPI4 or the closely related SaPI1028 (Fig. 6). One ET41 strain, RN5906, was also positive for type III integrase, suggesting that it contains a third SaPI, which would probably be related to SaPIm4 or SaPImw2 (see Fig. 3a). Of the six SaPI2 strains lacking *int*-I, two had a band hybridizing with the *int*-IV probe and two others had *int*-III. One of the former had two separate copies of *tst*, presumably one on each of its two SaPIs. These results are consistent with the clonality of ET41, but demonstrate at least one significant variant among the 11 ET41 strains studied. Of the other 17 strains, 10 had both SaPI2 junctional sequences, indicating occupation of the *att/int*-V site. Of these 10, three had the SaPI2 *ori*, five had *int*-V and two, 4455 and 4855, had neither. RN4455 had *int*-I, -III and -IV and presumably has a recombinant SaPI at the SaPI2 *att* site. RN4855 had a ~3.1 kb insert at the SaPI2 site, which probably represents the common SaPI remnant, SaPI6Δ. Of the final seven strains, three had neither of the SaPI2 junctions and four had $J_L$ only. We have not analysed

these further and suggest, provisionally, that they have either SaPI remnants or recombinants at the SaPI2 *att* site. Using primers homologous to *groEL*, which is present at the *att*-V site ($J_L$) and a putative Na transporter closely linked to the *att*-V right junction ($J_R$), we observed a 2.3 kb product with the same mobility as that obtained with MRSA252, which does not have any SaPI-like element at the *att*-V site, in two of the strains. This product includes a 1.4 kb segment that is adjacent to SaPI2 in 17 of the strains under study, including all of the ET41 strains, and contains four 'phage-related' PGs. It is very likely to be the result of horizontal transfer (from an unknown donor).

### Recombinational mosaicism

Four examples of SaPI evolution by probable recombination are SaPI2, SaPI4, SaPIN1 and SaPI1028. As shown in Fig. 6, SaPI2 and SaPI1028 have identical *rep/ori* modules but have different integrases and are located at different sites; SaPI4 and SaPI1028 have the same integrase and are located at the same site, but have divergent *rep/ori* modules; SaPI2 and SaPIN1 are not only at the same site and have the same regulatory and packaging modules, but have different *rep/ori* specificities and contain different accessory genes. Indeed, many of the strains listed in Table S2 (available with the online version of the journal) appear to have combinations of SaPI elements and, therefore, recombinant SaPIs. See for example, RN4455, which lacks SaPI2*int/ori* but has both SaPI2 junctions as well as types I, III and IV integrases, perhaps representing three different SaPIs, one of which is probably a recombinant inserted at the SaPI2 site. This raises the interesting question of how it may have integrated given that it lacks the cognate integrase; one possibility is that two SaPIs may have reciprocally recombined, with a crossover point between *int* and *att*, followed by insertion catalysed by the non-cognate integrase still present in the cell. If so, then one might expect to find the reciprocal recombinant in the same cell. Analysis of such recombination events and products is planned.
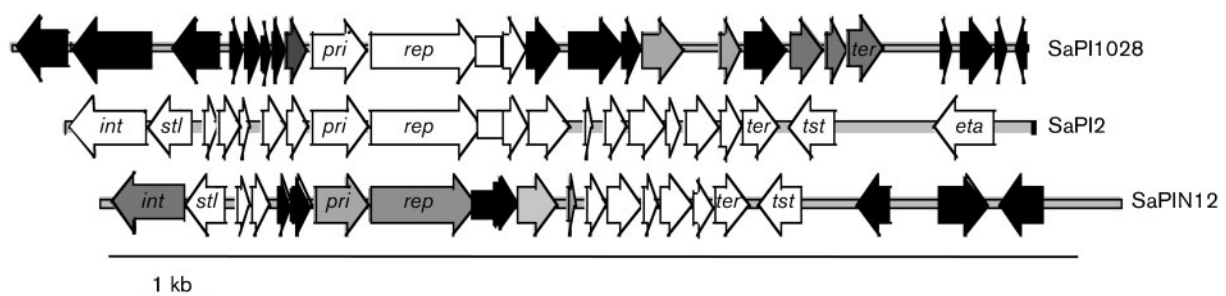


**Fig. 6.** Comparative maps of three different SaPIs. SaPI2 PGs are unshaded. Regions of SaPI1028 and SaPIN1 that share over 90 % nucleotide similarity with the corresponding regions of SaPI2 are also unshaded; these matching regions presumably represent recombinational exchanges. PGs of SaPI1028 and SaPIN1 that are between 50 and 90 % similar to SaPI2 are shaded in grey; dissimilar PGs are shaded in black. Gene labels are as described in Table S4 (available with the online version of this paper).

## Evolution

Four of the more recently described SaPIs, SaPI1028, SaPI4, SaPI122 and ShPI2, do not encode any of the known virulence factors or other accessory genes. Very possibly, these represent precursors of the superantigen-encoding mobile elements, as SaPI122 plus the nearly identical elements, SaPIn1 and SaPIm1, are all at the same site (See Table 1 and Fig. 4). It could be imagined that the acquisition of toxin genes occurred by recombination; Fig. S8 (available with the online version of the journal) illustrates the corresponding regions of SaPI2 and SaPI4, showing homologous sequences flanking *tst* in SaPI2 and SAR0084 in SaPI4. Recombination involving these sequences could have switched these two genes, generating a *tst*-containing SaPI4.

The *att*/*int*-V site is occupied, in five of the six sequenced strains by SaPI6Δs. Interestingly, there is a single base mismatch between the flanking *att* sites of all five of the SaPI6Δs, with one copy matching that of SaPIn1/m1, the other that of SaPI122 (Fig. S9, available with the online version of the journal), suggesting that their evolutionary history probably includes a recombination event between these. One other group V island, SaRIfusB, also has core sequence mismatches at the two junctions (Fig. S9).

Although all of the SaPIs are clearly co-ancestral, a fascinating question is that of how their highly conserved genome organization might have become associated with different insertion site specificities. The simplest idea is that the genome organization and insertion site specificities evolved independently and were later combined by recombination (see Fig. 6). Although their overall genomic organization is parallel to that of typical temperate phages, it seems clear that the SaPIs are not simply defective phages. Instead, we suggest that the SaPIs and the temperate phages have evolved in parallel from a common ancestor. Such a scheme would imply that the replication capability and regulatory organization evolved first and that encapsidation evolved later with the SaPIs 'learning' how to parasitize the phage encapsidation system to generate their small-headed particles. In this connection, it is interesting that SaPIs interfere profoundly with the phages that induce their ERP cycle but that this interference is at the level of encapsidation rather than replication (E. Maiques & others, unpublished).

## ACKNOWLEDGEMENTS

## REFERENCES

**Altemeier, W. A., Lewis, S. A., Schlievert, P. M., Bergdoll, M. S., Bjornson, H. S., Staneck, J. L. & Crass, B. A. (1982).** *Staphylococcus aureus* associated with toxic shock syndrome. *Ann Intern Med* **96**, 978–982.

**Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K. & other authors (2002).** Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**, 1819–1827.

**Bergdoll, M. S. & Schlievert, P. M. (1984).** Toxic shock syndrome toxin. *Lancet* **2**, 691.

**Campbell, A. M. (1969).** *Episomes.* New York: Harper & Row.

**Chu, M. C., Kreiswirth, B. N., Pattee, P. A., Novick, R. P., Melish, M. E. & James, J. F. (1988).** Association of toxic shock toxin-1 determinant with a heterologous insertion at multiple loci in the *Staphylococcus aureus* chromosome. *Infect Immun* **56**, 2702–2708.

**Diep, B. A., Gill, S. R., Chang, R. F., Phan, T., Chen, J., Davidson, M., Lin, F., Lin, J., Carleton, H. & Mongodin, E. (2006).** Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* **367**, 731–739.

**Fitzgerald, J. R., Monday, S. R., Foster, T. J., Bohach, G. A., Hartigan, P. J., Meaney, W. J. & Smyth, C. J. (2001).** Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *J Bacteriol* **183**, 63–70.

**Holden, M. T., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L. & other authors (2004).** Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A* **101**, 9786–9791.

**Ji, G., Beavis, R. & Novick, R. P. (1997).** Bacterial interference caused by autoinducing peptide variants. *Science* **276**, 2027–2030.

**Jones, C. L. & Khan, S. A. (1986).** Nucleotide sequence of the enterotoxin B gene from *Staphylococcus aureus*. *J Bacteriol* **166**, 29–33.

**Kreiswirth, B. N., O'Reilly, M. & Novick, R. P. (1984).** Genetic characterization and cloning of the toxic shock syndrome exotoxin. *Surv Synth Pathol Res* **3**, 73–82.

**Kreiswirth, B. N., Projan, S. J., Schlievert, P. M. & Novick, R. P. (1989).** Toxic shock syndrome toxin 1 is encoded by a variable genetic element. *Rev Infect Dis* **11** (Suppl. 1), S83–S88 (discussion S88–S89)

**Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K. & other authors (2001).** Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225–1240.

**Kuroda, M., Yamashita, A., Hirakawa, H., Kumano, M., Morikawa, K., Higashide, M., Maruyama, A., Inose, Y., Matoba, K. & other authors (2005).** Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. *Proc Natl Acad Sci U S A* **102**, 13272–13277.

**Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. (2005).** The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci U S A* **102**, 5174–5179.

**Lindsay, J. A., Kurepina, N. & Novick, R. P. (1997).** Clinical isolates of *Staphylococcus aureus* encode TSST-1 on genetic elements related to *S. aureus* pathogenicity island-1 (SaPI1). In *European Conference on Toxic Shock Syndrome*. Royal Society of Medicine, London, UK: Royal Society of Medicine.

**Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. (1998).** The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol* **29**, 527–543.

**Maiques, E., Úbeda, C., Tormo, M. A., Ferrer, M. D., Lasa, I., Novick, R. P. & Penadés, J. R. (2007).** Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *J Bacteriol* **189**, 5608–5616.

**Musser, J. M., Schlievert, P. M., Chow, A. W., Ewan, P., Kreiswirth, B. N., Rosdahl, V. T., Naidu, A. S., Witte, W. & Selander, R. K. (1990).** A single clone of *Staphylococcus aureus* causes the majority of cases of toxic shock syndrome. *Proc Natl Acad Sci U S A* **87**, 225–229.

Novick, R. P. & Subedi, A. (2007). The SaPIs: mobile pathogenicity islands of *Staphylococcus*. *Chem Immunol Allergy* **93**, 42–47.

Novick, R. P., Schlievert, P. & Ruzin, A. (2001). Pathogenicity and resistance islands of staphylococci. *Microbes Infect* **3**, 585–594.

O'Neill, A. J., Larsen, A. R., Skov, R., Henriksen, A. S. & Chopra, I. (2007). Characterization of the epidemic European fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *J Clin Microbiol* **45**, 1505–1510.

Ruzin, A., Lindsay, J. & Novick, R. P. (2001). Molecular genetics of SaPI1 – a mobile pathogenicity island in *Staphylococcus aureus*. *Mol Microbiol* **41**, 365–377.

Sato, H., Watanabe, T., Murata, Y., Ohtake, A., Nakamura, M., Aizawa, C., Saito, H. & Maehara, N. (1999). New exfoliative toxin produced by a plasmid-carrying strain of *Staphylococcus hyicus*. *Infect Immun* **67**, 4014–4018.

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M. & other authors (2005). Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *J Bacteriol* **187**, 7292–7308.

Todd, J., Fishaut, M., Kapral, F. & Welch, T. (1978). Toxic-shock syndrome associated with phage-group-I staphylococci. *Lancet* **2**, 1116–1118.

Ubeda, C., Tormo, M. A., Cucarella, C., Trotonda, P., Foster, T. J., Lasa, I. & Penades, J. R. (2003). Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol Microbiol* **49**, 193–210.

Ubeda, C., Maiques, E., Novick, R. P. & Penades, J. R. (2007a). SaPI operon I is required for SaPI packaging and is regulated by LexA. *Mol Microbiol* **65**, 41–50.

Ubeda, C., Penades, J. R. & Novick, R. P. (2007b). A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc Natl Acad Sci U S A* **65** ( (in press).

Ye, Z.-H. & Lee, C. (1993). Cloning, sequencing, and genetic characterization of regulatory genes, *rinA* and *rinB*, required for the activation of staphylococcal phage phi11 *int* expression. *J Bacteriol* **175**, 1095–1102.

Edited by: J. Parkhill