Rapid differentiation of *Staphylococcus aureus* isolates harbouring *egc* loci with pseudogenes *\(\psi\)ent1* and *\(\psi\)ent2* and the *selu* or *selu\(_v\)* gene using PCR-RFLP

Mark M. Collery and Cyril J. Smyth

Department of Microbiology, Moyne Institute of Preventive Medicine, Trinity College, University of Dublin, Dublin 2, Republic of Ireland

The *egc* locus of *Staphylococcus aureus* harbours two enterotoxin genes (*seg* and *sei*) and three enterotoxin-like genes (*selm*, *seln* and *selo*). Between the *sei* and *seln* genes are located two pseudogenes, *\(\psi\)ent1* and *\(\psi\)ent2*, or the *selu* or *selu\(_v\)* gene. While these two alternative *sei–seln* intergenic regions can be distinguished by PCR, to date, DNA sequencing has been the only confirmatory option because of the very high degree of sequence similarity between *egc* loci bearing the pseudogenes and the *selu* or *selu\(_v\)* gene. In *silico* restriction enzyme digestion of genomic regions encompassing the *egc* locus from the 3’ end of the *sei* gene through the 5’ first quarter of the *seln* gene allowed pseudogene- and *selu* or *selu\(_v\)*-bearing *egc* loci to be distinguished by PCR-RFLP. Experimental application of these findings demonstrated that endonuclease HindIII cleaved PCR amplimers bearing pseudogenes but not those with a *selu* or *selu\(_v\)* gene, while *selu* or *selu\(_v\)* -bearing amplimers were susceptible to cleavage by endonuclease HphI, but not by endonuclease HindIII. The restriction enzyme BccI cleaved *selu* or *selu\(_v\)*-harbouring amplimers at a unique restriction site created by their signature 15 bp insertion compared with pseudogene-bearing amplimers, thereby allowing distinction of these *egc* loci. PCR-RFLP analysis using these restriction enzymes provides a rapid, easy to interpret alternative to DNA sequencing for verification of PCR findings on the nature of an *egc* locus type, and can also be used for the primary identification of the intergenic *sei–seln* *egc* locus type.

**INTRODUCTION**

The staphylococcal enterotoxins (SEs) belong to the super-antigen family of exotoxins and are the only superantigens to possess emetic activity (Le Loir et al., 2003; Proft & Fraser, 2003). The International Nomenclature Committee for Staphylococcal Superantigens proposed that only staphylococcal superantigens that caused vomiting after oral administration in a primate model should be designated SEs, whereas other related toxins that either lacked emetic properties or had not been tested in a primate model should be designated staphylococcal enterotoxin-like toxins (SEls) (Lina et al., 2004). Fifteen years ago only five SEs were recognized, namely, SEA, SEB, SEC1/SEC2/SEC3, SED and SEE. Since then, through gene sequencing and partial and complete genome sequencing, 15 additional SEs and SEls have been described, namely, SEG–SEI, SEIJ–SEIR, SEIU, SEIU2 and SEIV, and three further variants of SEC (SEC-bovine, SEC-ovine and SEC-caprine) (reviewed by Le Loir et al., 2003; Smyth et al., 2004) (Thomas et al., 2006).

Following the identification of enterotoxins SEG and SEI (Munson et al., 1998), the genes encoding them were demonstrated to be part of a chromosomal operon, named the enterotoxin gene cluster (*egc*), comprising five genes, now designated *selo*, *selm*, *sei*, *seln* and *seg* in transcriptional order, and two pseudogenes, *\(\psi\)ent1* and *\(\psi\)ent2*, between the *sei* and *seln* genes (Jarraud et al., 2001a, b; Monday & Bohach, 2001). It was proposed that the *egc* cluster arose by gene duplication and variation from an ancestral gene through unequal crossing-over generated by recombination involving misalignment between non-allelic regions (Jarraud et al., 2001a). Subsequently, Letertre et al. (2003) have demonstrated that some *egc* clusters possess a novel gene between the *sei* and *seln* genes, designated *selu*, arising from a 15 bp insertion (5’-CTCTAAAATTTGATGG-3’) in the *\(\psi\)ent1* pseudogene sequence. Moreover, while the *selu* genes from three strains exhibit 99% nucleotide sequence
identity, one designated a variant (selu) yields a gene product with only 95–96 % amino acid sequence identity to the other three SElU proteins (Letertre et al., 2003).

Letertre et al. (2003) developed three sets of primers to allow distinction between strains possessing pseudogenes and the selu or selu gene in the egc locus. However, verification of PCR findings other than by DNA sequencing has not been described because of the very high sequence identity of the sei–seln intergenic region apart from the 15 bp insert. The development and application of a simple procedure for distinction of egc loci of Staphylococcus aureus bearing pseudogenes and the selu or selu gene is described herein. The method uses PCR-RFLP with either digestion by restriction endonucleases HindIII and HphI in tandem or endonuclease BcI digestion alone of PCR amplimers from the stop codon of the sei gene through the 5′ first quarter of the seln gene obtained using the PSE1 and PSE4 primers of Fitzgerald et al. (2003). The differential susceptibility or resistance of pseudogene- and selu- or selu+ -bearing amplimers allows distinction of these two types of egc loci.

**METHODS**

**Strains.** The fully genome-sequenced S. aureus strains RF122, Mu50 (ATCC 700699), MRSA 252, MW2, COL and NCTC 8325-4 were used as control strains of known egc locus status (Table 1). Ten isolates from chickens, goats, and cows, all of which were known to possess the egc locus, were also examined (Smyth et al., 2005). Eight bovine isolates of RAPD type 7 that were known to possess or lack the sec and tst genes (Fitzgerald et al., 2000), but which had not been included in the study of Smyth et al. (2005) and had not been further tested for the presence of SE and SEL genes, were also used (Table 1).

All isolates were confirmed as S. aureus on the basis of sugar fermentation tests (arabinose, maltose, mannitol, sucrose, trehalose,

### Table 1. Strains of S. aureus

<table>
<thead>
<tr>
<th>Strain/isolate</th>
<th>Intergenic sei–seln egc locus type*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF122</td>
<td>selu</td>
<td>Fitzgerald et al. (2001)</td>
</tr>
<tr>
<td>Mu50 (ATCC 700699)</td>
<td>ψent1 ψent2</td>
<td>Kuroda et al. (2001)</td>
</tr>
<tr>
<td>MRSA 252†</td>
<td>selu</td>
<td>Holden et al. (2004)</td>
</tr>
<tr>
<td>MW2†</td>
<td>egc−</td>
<td>Baba et al. (2002)</td>
</tr>
<tr>
<td>COL†</td>
<td>egc−</td>
<td>Dyke et al. (1966)</td>
</tr>
<tr>
<td>NCTC 8325-4†‡</td>
<td>egc−</td>
<td>Novick (1967)</td>
</tr>
<tr>
<td>A900322§</td>
<td>ψent1 ψent2</td>
<td>Jarraud et al. (2001a)</td>
</tr>
<tr>
<td>5l</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>12.11l</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>148II</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>1007II</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS37II</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS56§</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS59§</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS63§</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS27#</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>DS29#</td>
<td>egc+</td>
<td>Smyth et al. (2005)</td>
</tr>
<tr>
<td>46**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>83**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>88**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>92**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>100**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>101**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>109**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
<tr>
<td>112**</td>
<td>ND</td>
<td>Fitzgerald et al. (2000)</td>
</tr>
</tbody>
</table>

*egc+, possessing an egc locus; egc−, lacking an egc locus; ND, not determined.
†A derivative of a human isolate NCTC 8325 that is cured of prophages φ11, φ12 and φ13.
‡Bovine isolates, possessing the selo, selm, sei, seln and seg genes.
§Chicken isolates, possessing the selo, selm, sei, seln and seg genes.
#Goat isolates, possessing the selo, selm, sei, seln and seg genes.
**All isolates of bovine origin of RAPD type 7 (Fitzgerald et al., 2000).
In silico analyses of sei–seln intergenic regions of egc loci.
Fasta files of DNA sequences of egc loci, sei–seg intergenic regions, and selu and selu genes were retrieved from the National Centre for Biotechnology Information (NCBI) nucleotide database (Table 2). For strains A900322, FR1572, RF122, 382F, Mu50, N315 and MRSA 252, the DNA sequences of the forward primer PSE1 and the reverse primer PSE4 of Letertre et al. (2003) were used to identify the sei–seln intergenic regions bearing pseudogene yent1 and yent2 and the selu or selu genes. The ORF Finder (NCBI) was used to identify the pseudogenes yent1 and yent2 and the selu or selu genes (http://www.ncbi.nlm.nih.gov/projects/gorf/). Alignments performed using the CLC Free Workbench software version 3.0.2 (CLC bio; http://www.clcbio.com) on these PSE1–PSE4 sequences allowed identification of the PSE2 primer sequence of Letertre et al. (2003) in the selu and selu genes and of the 15 bp insertion that converts the pseudogene to the selu or selu gene. Unless otherwise indicated, the nucleotide numbering system used herein refers to the first base of the PSE1 forward primer (T) as nt 1.

The database sequences for the selu genes of strains 352E, 383F and FR1137 were incomplete with respect to the PSE1–PSE4 region. Using the Blast program, the available sequences for strains 352E (AY205305), 383F (AY205307) and FR1137 (AY205306) gave best matches of 1040/1052 (98.9%), 1041/1052 (99.0%) and 1024/1024 (100%), respectively, with that of strain RF122, in line with their own inter-identity values (Letertre et al., 2003). Accordingly, the necessary sequence additions to the 5' and/or 3' ends to obtain complete in silico PSE1–PSE4 regions for these three strains were based on the complete PSE1–PSE4 sequence of strain RF122. All of the complete or completed PSE1–PSE4 sequences were analysed using RestrictionMapper version 3 (http://www.restrictionmapper.org) to identify restriction enzyme cleavage sites.

DNA preparation. Extraction of genomic DNA from S. aureus was performed as previously described, with some minor modifications (Fitzgerald et al., 1997). The procedure involves lysostaphin lysis, proteinase K treatment, addition of EDTA, sarkosyl and cetyltrimethylammonium bromide (CTAB), extraction with chloroform: isoamyl alcohol and phenol/chloroform: isoamyl alcohol, and precipitation with ethanol.

PCR. All primers were synthesized by MWG Biotech. For amplification of the sei–selu intergenic region bearing the yent1 and yent2 pseudogenes and the selu or selu gene, the forward/reverse primer pairs PSE1/PSE4, PSE2/PSE4 and PSE2/PSE6 described by Letertre et al. (2003) were used. The primers PSE1 and PSE4 are designed to amplify the egc locus region encoding either the pseudogenes yent1 and yent2 or the selu gene from the stop codon of the sei gene through the first 172 nt of the selu gene. The 27-mer forward primer PSE2 includes the specific 15 nt insertion of the selu or selu gene. PCR assays with the primer pairs PSE2/PSE4 and PSE2/PSE6 yield 790 and 142 bp amplimers, respectively, when the selu or selu gene is present (Letertre et al., 2003).

PCR was performed in a final volume of 25 μl. The reaction mix contained 2–2.5 pmol each primer. Each reaction contained 15.2 μl H2O, 1 μl each primer (forward/reverse combinations PSE1/PSE4, PSE2/PSE4, PSE2/PSE6), 1 μl dNTP mix, 0.3 μl Taq polymerase (5 U μl⁻¹), 2.5 μl 10 × Thermophilic DNA polymerase buffer (Mg²⁺ free), 3 μl MgCl₂ (15 mM) and 1 μl template DNA. All reagents were supplied by Promega. Thermal cycling conditions were: 94°C for 4 min, 7 cycles of (94°C for 30 s, 56°C for 30 s, 72°C for 4 min), 21 cycles of (94°C for 30 s, 56°C for 35 s, 72°C for 4 min) and a final elongation step of 72°C for 7 min. DNA preparations from strains A900322, Mu50, RF122, MRSA 252, MW2 and NCTC 8325-4 were used as controls.

To control for the presence of sufficient template DNA, the test samples were examined by PCR with a primer set that anneals to staphylococcal 16S rRNA genes, generating a 228 bp amplimer (Monday & Bohach, 1999), using the thermal cycling conditions of Smyth et al. (2005).

Table 2. DNA sequences of egc loci, sei–seg intergenic regions, and selu and selu genes

<table>
<thead>
<tr>
<th>Strain</th>
<th>Origin and egc locus type</th>
<th>Accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A900322</td>
<td>Tampon – toxic shock syndrome; yent1 yent2</td>
<td>AF285760</td>
<td>Jarraud et al. (2001a)</td>
</tr>
<tr>
<td>FR1572</td>
<td>Nasal isolate, Easter Island; yent1 yent2</td>
<td>AF156894</td>
<td>Monday &amp; Bohach (2001)</td>
</tr>
<tr>
<td>Mu50†</td>
<td>Surgical wound, VRSAb (1997); yent1 yent2</td>
<td>BA000017</td>
<td>Kuroda et al. (2001)</td>
</tr>
<tr>
<td>N315†</td>
<td>Pharyngeal smear, MRSAa (1982); yent1 yent2</td>
<td>BA000018</td>
<td>Kuroda et al. (2001)</td>
</tr>
<tr>
<td>RF122</td>
<td>SaPlblov bovine isolate (1993); selu</td>
<td>AJ938182</td>
<td>Fitzgerald et al. (2001)</td>
</tr>
<tr>
<td>383F</td>
<td>Isolate from rice salad (1987); selu</td>
<td>AY205306</td>
<td>Letertre et al. (2003)</td>
</tr>
<tr>
<td>352E</td>
<td>Isolate from fromage frais (1981); selu</td>
<td>AY205305</td>
<td>Letertre et al. (2003)</td>
</tr>
<tr>
<td>FR1137</td>
<td>Isolate from thgh abscess (1933); selu</td>
<td>AY205307</td>
<td>Bergdoll et al. (1965)</td>
</tr>
<tr>
<td>382F</td>
<td>Isolate from water (1988); selu</td>
<td>AY158703</td>
<td>Letertre et al. (2003)</td>
</tr>
<tr>
<td>MRSA 252†</td>
<td>Postoperative infection EMRSA-16 clone‡ (1997); selu</td>
<td>BX371856</td>
<td>Holden et al. (2004)</td>
</tr>
</tbody>
</table>

*A C. L. Wong, Food Research Institute, University of Wisconsin, Madison, personal communication.
†Genome-sequenced human isolates.
‡MRSA, meticillin-resistant S. aureus; VRSA, vancomycin-resistant S. aureus; EMRSA, epidemic MRSA.
PCR products were analysed on 2% (w/v) agarose gels alongside a 100 bp ladder (NEB), stained with ethidium bromide and visualized under UV light.

**Restriction enzyme digestion of PSE1–PSE4 amplimers.** Using the above PCR protocol, forward/reverse primer pair PSE1/PSE4 was used to amplify the pseudogene- or selu/selu-, containing region. PCR product (5 μl) was run on a 1% agarose gel to confirm that the reaction had been successful. Restriction endonuclease digestion was performed in a final volume of 30 μl. For endonuclease HindIII cleavage, this contained 17 μl H2O, 6 μl PCR product, 3 μl 10× Buffer E (60 μM Tris-HCl, pH 7.5, 1 M NaCl, 60 mM MgCl2, 10 mM DTT), 3 μl 10× BSA (100 μg ml−1) and 1 μl HindIII restriction endonuclease (Promega; 10 U μl−1). The digestion mixture was incubated overnight at 37°C. Endonuclease Hphl cleavage was carried out as above, except that the digestion mixture contained 3 μl 10× NE Buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) and 1 μl Hphl restriction endonuclease (NEB; 5 U μl−1). For endonuclease Bcll, the digestion mixture contained 3 μl 10× NE Buffer 1 (10 mM Bistris propane-HCl, 10 mM MgCl2, 1 mM DTT, pH 7.0) and 1 μl Bcll restriction endonuclease (NEB; 10 U μl−1). All digests were analysed on 2.5% (w/v) agarose gels alongside a ladder (NEB).

**Sequencing of PSE1–PSE4 PCR product.** PCR was performed for the PSE1–PSE4 amplifier of strain A900322. The PCR product was purified using the High Pure PCR Product Purification kit (Roche) according to the manufacturer’s instructions. The PCR amplifier was sequenced by GATC Biotech. The derived sequence (accession no. DQ993159) was aligned with the complete egc locus sequence of strain A900322 (AF285760) and with the seq-selg intergenic region of strain FR1572 (AF156894) using the CLC Free Workbench software version 3.0.2. The BLAST program (http://www.ncbi.nlm.nih.gov/blast/) was used to interrogate the NCBI database for matches for the derived sequence.

**RESULTS**

**In silico analyses of PSE1–PSE4, seq–selg intergenic regions of egc loci**

Alignment of the PSE1–PSE4 region sequences of the four ψent1+ ψent2+ strains A900322 (1066 nt), and Mu50, N315 and FR1572 (each 1135 nt), revealed a 69 nt deletion in the ψent1–ψent2 sequence of strain A900322 between nt G3462 and nt G3463 of the original egc cluster sequence AF285760 compared with those of the other three ψent1+ ψent2+ strains in which this 69 nt sequence is identical (5′-ACCGAGCATGATGGAAATCAAATAGATAAAAAT-ATTCAACTGATAACTCTCATATAATTTAATTTAA-3′) (Fig. 1). This 69 nt sequence is also 100% conserved in the sequences of two of the selu+ strains (383F, 352E) and the other two selu+ strains (382F and MRSA252) compared. In the other two selu+ strains RF122 and FR137 there is an inversion of TC to CT at nt 37 and nt 38 of the 69 nt sequence (see TC bold type in 69 nt sequence above). The missing 69 nt in the strain A900322 sequence occur in the 23 nt ψent1 and ψent2 reading-frame overlap region that was originally described by Jarraud et al. (2001a) to harbour the start codon of pseudogene ψent2 and the stop codon of pseudogene ψent1, namely, 5′-3448-ATG[ψent2]TATGGGCGGTGGTG[A69 nt]GTITATG[A70 nt]AG [ψent1]-3470 3′ (bold type signifies features in the sequence mentioned above).

In strain A900322, the ψent1–ψent2 pseudogene sequence starts at nt 154 and ends at nt 856 of the PSE1–PSE4 amplifier sequence, whereas in strains Mu50, N315 and FR1572, the ψent1–ψent2 sequence starts at nt 154 and ends at nt 925.

The PSE1–PSE4 sequences of three of the selu+ (RF122, 352E, FR137) and the two selu+ (382F, MRSA252) strains are 1149 nt long. That of strain 383F (selu−) is 1 nt longer due to the insertion of a thymine in its PSE1–PSE4 sequence outside the selu ORF between three thymines (T68–T70). In PSE1–PSE4 sequences, the selu or selu+ gene begins at nt 154 (nt 155 in strain 383F) and ends at nt 939 (940 in strain 383F). Analysis of the 1149/1150 nt long selu and selu+ sequences compared with PSE1–PSE4 sequences of strains A900322, Mu50, N315 and FR1572 revealed the 15 nt insertion between G218 and T219 of the ψent1 sequences (5′-219-CTCTAAAATTGATGG-233 3′ of the selu or selu+ genes) [this 15 nt sequence comprises nt 13–27 of the PSE2 primer of Letertre et al. (2003)]. All six of the selu- or selu+ bearing sequences have a single adenine deletion compared with the ψent1 sequences of strains A900322, Mu50, N315 and FR1572 from a run of seven adenines (A365–A371 of the ψent1 pseudogenes corresponding to nt A360–A385 of the selu and selu+ genes). This single adenine deletion abolishes the opal/amber stop codon TGA present at the end of pseudogene ψent1.

![Fig. 1](http://jmm.sgmjournals.org)
Table 3. In silico analysis of restriction endonuclease cleavage of sei–seln intergenic sequences of the egc locus between primers PSE1 and PSE4 containing pseudogenes \( \psi \text{ent1} \) and \( \psi \text{ent2} \) or the selu or selu\(_v\) gene

<table>
<thead>
<tr>
<th>Restriction endonuclease</th>
<th>Sizes (bp) of restriction fragments of PSE1–PSE4 regions for indicated strains*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A900322 (1066 bp, ( \psi \text{ent1} \ \psi \text{ent2} ))</td>
</tr>
<tr>
<td>HindIII</td>
<td>889</td>
</tr>
<tr>
<td>HphII</td>
<td>176</td>
</tr>
<tr>
<td>Bcl</td>
<td>569</td>
</tr>
<tr>
<td></td>
<td>496</td>
</tr>
<tr>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>

*See Table 2 for accession numbers of sequences.

\(^\dagger\)Second value is that for restriction fragments of strain 383F, the PSE1–PSE4 amplimer of which is 1 nt longer due to the insertion of a thymine outside the selu ORF between three thymines (T 68–T 70).

\(^\ddagger\)Deletion of a single adenine residue that abrogates the stop codon TGA of pseudogene \( \psi \text{ent1} \) leads to this fragment of the selu and selu\(_v\) genes being 1 nt shorter.

Each of the 10 PSE1–PSE4 sequences was analysed using the RestrictionMapper version 3 program (Table 3). Of the enzymes with one restriction site, endonuclease HindIII cut the PSE1–PSE4 regions between A 958 and A 959 in strains Mu50, N315 and FR1572 (Fig. 1), A 889 and A 890 in strain A900322, 33 nt downstream of the \( \psi \text{ent2} \) stop codon TAA and 5 nt upstream of the start codon of the selu gene in the strains with pseudogenes \( \psi \text{ent1} \) and \( \psi \text{ent2} \), but did not cut PSE1–PSE4 regions in strains bearing the selu and selu\(_v\) genes (Table 3). In contrast, restriction endonuclease HphII cut the PSE1–PSE4 region between T 850 and A 851 in strains with selu or selu\(_v\) genes (Fig. 1), 86 nt upstream of the stop codon, but did not cut PSE1–PSE4 regions in strains bearing the pseudogenes \( \psi \text{ent1} \) and \( \psi \text{ent2} \).

The only enzyme found to cut the sequence of the 27-mer PSE2 forward primer (Lerette et al., 2003) was restriction endonuclease Bcl. The insertion of the 15 bp sequence creates a unique Bcl cleavage site between A 376 and A 377 (nt 17 and nt 18 of the PSE2 primer) of PSE1–PSE4 amplimers of selu\(^+\) and selu\(_v\)\(^+\) strains (Fig. 1). PSE1–PSE4 regions bearing pseudogenes \( \psi \text{ent1} \) and \( \psi \text{ent2} \) are cleaved into three fragments, whereas those bearing selu and selu\(_v\) genes are cleaved into four fragments (Table 3). Pseudogene- and selu- or selu\(_v\)-bearing strains can be readily distinguished by the presence of a 496 nt Bcl fragment and 376/377 nt and 135 nt Bcl fragments, respectively (Table 3).

PCR-RFLP analysis of PSE1–PSE4 amplimers using restriction endonucleases HindIII and HphII in tandem

Strains Mu50 and A900322 (\( \psi \text{ent1}^+ \ \psi \text{ent2}^+ \)), RF122 (selu\(^+\)) and MRSA 252 (selu\(^+\)) were examined using the PSE1/PSE4, PSE2/PSE4 and PSE2/PSE6 primer sets of Lerette et al. (2003) (Fig. 2a–c). All four strains yielded PCR products of approximately 1135–1150 bp with the PSE1/PSE4 primers. Strains Mu50 and A900322 did not give amplifiers with the PSE2/PSE4 and PSE2/PSE6 primer pairs, confirming their \( \psi \text{ent1}^+ \ \psi \text{ent2}^+ \) status. In contrast, strains RF122 and MRSA 252 produced PCR products of approximately 790 and 142 bp with primers PSE2/PSE4 and PSE2/PSE6, respectively (Fig. 2b, c), confirming the presence of selu and selu\(_v\) genes, respectively. Strains NCTC 8325-4, MW2 and COL were subjected to PCR as negative controls for the egc locus. None yielded PSE1/PSE4, PSE2/ PSE4 or PSE2/PSE6 amplimers.

To test whether isolates bearing either the pseudogenes \( \psi \text{ent1} \) and \( \psi \text{ent2} \) or the selu gene or the selu\(_v\) gene could be distinguished using restriction endonucleases HindIII and HphII, the PSE1–PSE4 amplimers of strains A900322, Mu50, RF122 and MRSA 252 were incubated overnight with these restriction enzymes. As anticipated from the in silico analysis, amplimers of strains bearing the pseudogenes (A900322 and Mu50) were digested by endonuclease HindIII, but not by endonuclease HphII, whereas amplimers of strains harbouring the selu or selu\(_v\) gene (RF122 and MRSA 252) were restricted by endonuclease HphII, but not by endonuclease HindIII, yielding fragments of the expected approximate sizes (Fig. 2d, e).

Eight bovine isolates of RAPD type 7 (Fitzgerald et al., 2000), which were known to possess or lack the sec and tsr genes but had not been otherwise screened for SE or SEI genes, were tested for the presence of pseudogenes or the selu/selu\(_v\) gene by PCR and PCR-RFLP analysis. In all cases PSE1–PSE4, PSE2–PSE4 and PSE2–PSE6 amplimers characteristic of isolates possessing an egc locus with the selu or selu\(_v\) gene were obtained. These isolates were all confirmed to have the
**selu or selu\textsubscript{v} gene** on the basis of the \textit{Hph\textsubscript{I}} susceptibility and \textit{Hind\textsubscript{III}} resistance of their PSE1–PSE4 amplimers (data not shown).

**PCR-RFLP analysis of PSE1–PSE4 amplimers using restriction endonuclease \textit{Bcc\textsubscript{I}}**

PCR-RFLP analysis of PSE1–PSE4 amplimers using endonuclease \textit{Bcc\textsubscript{I}} was carried out for the \textit{egc} archetypal strain A900322 and the genome-sequenced strains Mu50, RF122 and MRSA 252 (Fig. 2f). The pseudogene-positive strains A900322 and Mu50 yielded a restriction product of approximately 496 bp, whereas strains RF122 and MRSA 252, with \textit{selu} and \textit{selu\textsubscript{v}} genes, respectively, yielded restriction products of approximately 376 and 135 bp (Fig. 2f, Table 3).

Ten animal isolates known to possess an \textit{egc} locus (the \textit{selo}, \textit{selm}, \textit{sei}, \textit{seln} and \textit{seg} genes) – five from cows, two from goats and three from chickens – were identified as having pseudogenes and the \textit{selu} or \textit{selu\textsubscript{v}} gene using PCR with the PSE primer sets and PCR-RFLP with endonucleases \textit{Hind\textsubscript{III}} and \textit{Hph\textsubscript{I}} in tandem. The three chicken isolates and two of the bovine isolates (DS37 and 1007) possessed pseudogenes, while the two goat isolates and three of the bovine isolates had the \textit{selu} or \textit{selu\textsubscript{v}} gene (Table 1). Their PSE1–PSE4 amplimers were then examined by PCR-RFLP using restriction endonuclease \textit{Bcc\textsubscript{I}} in an operator-blinded manner. The \textit{egc} loci were all correctly identified, those bearing pseudogenes giving a restriction product of approximately 496 bp and those bearing a \textit{selu} or \textit{selu\textsubscript{v}} gene giving restriction products of approximately 376 and 135 bp, respectively (data not shown).

**Sequencing of the PSE1–PSE4 amplifier of strain A900322**

There was no evidence from the PCR-RFLP studies to suggest that the \textit{egc} locus of strain A900322 differed from that of the \textit{\textit{\psi\textsubscript{ent1}}\textsuperscript{+} \textit{\textit{\psi\textsubscript{ent2}}\textsuperscript{+}}} strain Mu50. Given that the missing 69 nt were present in all other published sequences of \textit{selu} or \textit{selu\textsubscript{v}} genes and \textit{\textit{\psi\textsubscript{ent1}}\textsuperscript{+} \textit{\textit{\psi\textsubscript{ent2}}\textsuperscript{+}}} pseudogenes in the NCBI database, resequencing of the pseudogene region of the \textit{egc} locus of strain A900322 was performed. The 694 nt-long sequence obtained was from nt A 64 to nt G 757 of the...
PSE1–PSE4 amplimer (accession no. DQ993159). This encompasses all of pseudogene \textit{ψent1} and the first 288 nt of pseudogene \textit{ψent2}, including the crucial pseudogene \textit{ψent1–ψent2} overlap region in which the 69 nt sequence in question would occur, if present.

Using the CLC Free Workbench software version 3.0.2, this 694 nt sequence was aligned with the complete egc locus sequence of strain A900322 (AF285760; 6418 nt) and the \textit{sei–seg} intergenic spacer region of strain FR1572 (AF156894; 2016 nt). With the exception of 1 nt difference (C 342 in AF156894 versus T 189 in DQ993159), the derived sequence showed complete nucleotide concordance with nt A 154–G 847 of the strain FR1572 intergenic spacer region of Monday & Bohach (2001). In the case of the complete egc locus sequence of Jarraud \textit{et al.} (2001a), there was a 100 % match between the derived sequence DQ993159 and the complete egc locus sequence AF285760 of strain A900322 from nt A 2982 to nt G 3606, barring the absence of the 69 nt between nt G 3462 and nt G 3463 of the AF285760 sequence. Using \textsc{BLAST}, the 694 nt DQ993159 sequence gave a 100 % match with \textit{ψent1–ψent2} regions of the genomic sequences of strains Mu50 and N315, both of which possess the 69 nt sequence.

Taken together, these findings revealed that the missing 69 nt were present in the resequenced pseudogene \textit{ψent1–ψent2} overlap region of strain A900322. The 23 nt \textit{ψent1} and \textit{ψent2} reading-frame overlap region of strain A900322 originally described by Jarraud \textit{et al.} (2001a) to harbour the start codon of pseudogene \textit{ψent2} and the stop codon of pseudogene \textit{ψent1} is thus 26 nt long, as it is in the other \textit{ψent1}+ \textit{ψent2}+ strains Mu50, N315 and FR1572.

**DISCUSSION**

To date, confirmation of PCR findings in relation to the nature of the egc locus has only been possible by sequencing of PCR products. As demonstrated herein, PCR-RFLP can be utilized for verification of the nature of egc loci determined by PCR when applying the primer sets of Letertre \textit{et al.} (2003). Chini \textit{et al.} (2006) have recently used restriction endonuclease \textit{HindIII} to distinguish MRSA strains bearing the egc1 locus (\textit{selo selm seli ψent1 ψent2 seln seg}) and the egc2 locus (\textit{selo selm seli selu selu seg}). Using a clearly defined amplimer, the present findings not only confirm the value of using endonuclease \textit{HindIII}, but further extend the use of restriction enzymes by inclusion of endonuclease \textit{Hphi} in tandem for reverse susceptibility of the egc1 and egc2 loci. Furthermore, the usefulness of endonuclease \textit{Bcl} is pinpointed by the fact that the 15 bp insert leading to the formation of the \textit{selu} and \textit{selu}+ genes from pseudogenes \textit{ψent1} and \textit{ψent2} creates a unique cleavage site. PCR-RFLP either with endonucleases \textit{HindIII} and \textit{Hphi} in tandem or with endonuclease \textit{Bcl} alone with the PSE1–PSE4 amplimer can also be used as a primary screening test for these egc loci. The described PCR-RFLP procedure is rapid and amenable to the screening of batches of clinical isolates. However, strains with atypical egc loci are known to occur (Blaiotta \textit{et al.}, 2004; Thomas \textit{et al.}, 2006). Both the PCR procedure of Letertre \textit{et al.} (2003) and the PCR-RFLP procedure described herein may be affected by insertion sequences and by egc gene recombination or deletion events such as those reported by Thomas \textit{et al.} (2006).

Direct repeats ATTT, AAGG and CATGAT were identified as the insertion sites of the newly described transposase 8-like/retrovirus integrase-like cassette in the \textit{sei, seln} and \textit{seg} genes of isolates apparently lacking these genes by PCR analysis (Thomas \textit{et al.}, 2006). Analysis of the \textit{sei–seln} intergenic region reveals that the nucleotide motif ATTT occurs 21 times in strain Mu50 and 22 times in strains RF122 and MRSA 252 (equivalent to 21 % of the total ATTT direct repeats in the egc locus of archetypal strain A900322). Thus, on this basis the \textit{sei–seln} intergenic region would seem to be a potential hot spot for insertion sequences such as the transposase 8-like/retrovirus integrase-like cassette.

The \textit{in silico} analysis of egc loci based on the available NCBI database sequences annotated for the first time that the \textit{sei–seln} intergenic region of the archetypal egc+ strain A900322 (AF285760; Jarraud \textit{et al.}, 2001a) differed from the inter-pseudogenic sequences of three other \textit{ψent1}+ \textit{ψent2}+ strains, one of which (strain FR1572) is the egc prototype strain of Monday & Bohach (2001), by the absence of a 69 nt-long stretch, as well as from the sequences of six \textit{selu}+ or \textit{selu}+ strains. This strongly suggested either that the \textit{sei–seln} intergenic region of strain A900322 was atypical due to a unique deletion or that the sequence lodged in the database was incorrect or faulty. Resequencing (DQ993159) and sequence alignments confirmed the presence of the missing 69 nt in the \textit{ψent1–ψent2} overlap region of strain A900322.

Thomas \textit{et al.} (2006) have reported a new \textit{sei–seln} intergenic gene in \textit{S. aureus} strain A900624 that they designate \textit{selu2}. The \textit{selu2} gene results from a single adenine deletion which converts the 772 nt \textit{ψent1–ψent2} region into a 771 nt ORF. However, as pointed out herein, this same single adenine deletion is present in all the \textit{selu} and \textit{selu}+ gene sequences in the NCBI database. Moreover, the \textit{selu} 2 ORF differs in crucial respects from all published \textit{selu} and \textit{selu}+ sequences. The latter are 786 nt long and possess the 15 nt signature sequence 5’-219-CTCTAAATTGATGG-233 3’ between nt G 218 and nt T 219 of pseudogene \textit{ψent1}. From the description of the \textit{selu2} gene, namely, that nt 1 to nt 402 demonstrated 99.3 % (399/402) identity with pseudogene \textit{ψent1} of strain Mu50, and that nt 376 to nt 771 had 100 % identity to pseudogene \textit{ψent2} of strain Mu50, the \textit{selu2} gene is not a \textit{selu} or \textit{selu}+ gene \textit{sensu stricto} but merely a 1 nt frameshift of the \textit{ψent1–ψent2} sequence that deletes the opal/umber stop codon of pseudogene \textit{ψent1}.

Furthermore, while the nature of three nucleotide differences in the first 402 nt of the \textit{selu}2 gene are not available through the NCBI database, use of RestrictionMapper version 3 in \textit{silico} with the PSE1–PSE4
sequence of strain Mu50 with the single adenine removed confirmed that the selu2 gene is cleaved by endonuclease HindIII, but not by endonuclease HphI, and by endonuclease BclI to yield a 496 bp fragment. It is thus not possible to distinguish selu2-bearing from ψent1–ψent2-bearing isolates using the PCR-RFLP procedures described herein. Since the selu2 ORF is distinctly different from all selu genes described to date in the absence of the signature 15 nt descriptor of selu genes and its endonuclease HindIII, HphI and Bcl cleavage characteristics, we propose that this gene be redesignated the selu gene. This is coincidentally appropriate, since the single adenine deletion from the ψent1–ψent2 sequence removes the sixth-from-last tryptophan residue of the translated ψent1 sequences of strains A900322, Mu50, N315 and FRI572.

Neither the PCR-RFLP analyses herein nor the procedure of Chini et al. (2006) distinguishes selu- from selu+-bearing isolates. However, this appears to be possible using the PCR-RFLP scheme of Blaiotta et al. (2004, 2006). These authors demonstrated seven different PCR-RFLP groups based on restriction of a 3375 bp seln–seg amplimer using endonucleases EcoRI, TaqI, Alul and CfoI. Strain A900322 of the egc1 locus type belonged to REA group 2, strain FR1137 of the egc2 locus type belonged to REA group 2, and strain AB8802, probably of the egc3 locus type because it possesses the sei and seg genes that co-exist with the selu gene in strain 382F, belonged to REA group 6. However, these PCR-RFLP differences are not based on differences in the sei–seln intergenic regions of egc loci, as endonucleases TaqI, Alul and CfoI do not cut the PSE1–PSE4 amplimer, and endonuclease EcoRI has the same single cleavage site on the basis of the restriction analyses carried out herein using RestrictionMapper version 3, but rather are dependent on the presence or absence of sequence variation in other genes, e.g. sei, seg, and selu.

While there has been a plethora of reports on the occurrence of the five common genes of the egc locus (selo selm sei seln seg) in strains of S. aureus of human and animal origin over the past 5 years, very limited data exist on the relative causing potential of S. aureus harbouring these different intergenic sei–seln egc loci types.

ACKNOWLEDGEMENTS

M. M. C. was in receipt of a studentship from the Sarah Purser Medical Research Fund and of a Trinity College postgraduate research studentship. Thanks to Amy Wong and Marie-Laure DeBuyser, who provided information on the origins of Food Research Institute (FRI), University of Wisconsin, and Agence Française de Sécurité des Aliments (AFSSA) isolates, to Tim Foster and Sophie Jarraud, who provided control strains of known egc type, and to Davida Smyth and Ross Fitzgerald, who originally characterized the animal isolates.

REFERENCES


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.

Le Loir, Y., Baron, F. & Gautier, M. (2003). Staphylococcus aureus and
food poisoning. Genet Mol Res 1, 63–76.

of a new putative enterotoxin SEU encoded by the egc cluster of

Lina, G., Bohach, G. A., Nair, S. P., Hiramatsu, K., Jouvin-Marche, E.

detection of genes for Staphylococcus aureus enterotoxins, exfoliative
toxins, toxic shock syndrome toxin 1, and methicillin resistance.

detect classical and newly described pyrogenic exotoxin genes in

enterotoxins G and I are linked and separated by DNA related to other

Identification and characterization of staphylococcal enterotoxins types

Novick, R. P. (1967). Properties of cryptic high-frequency transdu-
cing phages in Staphylococcus aureus. Virolology 33, 156–166.


Schneider, T., Senn, M. M., Berger-Bächli, B., Tossi, A., Sahli, H.-G.
& Wiedemann, I. (2004). In vitro assembly of a complete, pentaglycine
interpeptide bridge containing cell wall precursor (lipid II-Gly5) of

(2004). Staphylococcus aureus: from man or animals – an enterotoxin
iceberg? In Food Pathogen Epidemiology: Microbes, Maladies and
Methods, Proceedings of European Union Risk Analysis Information
Network (EU-RAIN) Conference, Padua, Italy, pp. 85–102. Edited
by B. Maunsell, J. Sheridan & D. J. Bolton. Dublin Teagasc: The
National Food Centre.

Smyth, D. S., Hartigan, P. J., Meaney, W. J., Fitzgerald, J. R.,
genesis encoded by the egc cluster and SaPbov are predominant
among Staphylococcus aureus isolates from cows, goats, sheep, rabbits

Thomas, D. Y., Jarraud, S., Lemercier, B., Cozon, G., Echasserieau, K.,
Staphylococcal enterotoxin-like toxins U2 and V, two new staphy-
lococcal superantigens arising from recombination within the