Mapping the genetic diversity within major clonal complexes of meticillin-resistant *Staphylococcus aureus* utilizing genome-wide fluorescent amplified fragment length polymorphism markers

Darshana Desai,1 Matthew J. Ellington,2 Catherine Arnold1 and Meeta Desai1

Correspondence
Meeta Desai
meeta.desai@hpa.org.uk

1Microbiology Services Division Colindale, Health Protection Agency – Department for Bioanalysis and Horizon Technologies, London NW9 5EQ, UK
2Microbiology Services Division, Cambridge Laboratory, Health Protection Agency, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 00W, UK

The genetic diversity between major meticillin-resistant *Staphylococcus aureus* (MRSA) lineages was probed using fluorescent amplified fragment length polymorphism (FAFLP) as a random genome sampling tool. Genomic DNA was digested with endonucleases *Bgl*II and *Csp*6I and a subset of the restricted fragments were amplified using the primer pair *Bgl*II + A and *Csp*6I + 0. Sixty-seven FAFLP profiles consisting of 46–68 amplified fragments ranging in size from 50 to 600 bp were exhibited amongst the 71 isolates analysed. Cluster analysis of FAFLP data revealed concordance with *spa* typing and MLST clonal complexes (CC), with isolates of each CC grouping in the same FAFLP cluster. Furthermore, FAFLP could differentiate subtypes within the homogeneous CC22 isolates and also between MLST sequence types 8 and 239. The discriminatory power of FAFLP was 0.998 compared to values of 0.975 and 0.909 for *spa* typing and MLST, respectively. Thus, FAFLP analysis proved to be a rapid, reproducible and high-resolution tool that displayed the microheterogeneity within MRSA lineages. Using FAFLP data, lineage-specific fragments were identified and sequenced; these encoded toxins, antibiotic resistance determinants and bacteriophage resistance factors. Lineage-specific sequence variations were observed, which may provide insights into the evolution and fitness of successful lineages. This will also aid in the development of rapid and high-throughput diagnostic PCR-based assays for the identification of MRSA lineages in resource-poor settings.

INTRODUCTION

Meticillin-resistant *Staphylococcus aureus* (MRSA) has spread globally as a nosocomial pathogen and, in more recent years, in the community (Zetola *et al.*, 2005). It is associated with high rates of morbidity and mortality and, as such, represents a significant clinical burden (Cosgrove *et al.*, 2003; Yoon *et al.*, 2005). In the UK, distinct MRSA lineages, as defined by multilocus sequence typing (MLST), have become successful and pandemic. Molecular laboratory tools which highlight the genetic diversity within MRSA lineages aid in the surveillance and understanding of the evolution and population structure of MRSA, as well as the development of effective infection control measures (Hunter & Gaston, 1988).

Numerous molecular methods have been used to differentiate between lineages of MRSA such as MLST, staphylococcal cassette chromosome *mec* (*SCCmec*) typing, *spa* typing and amplified fragment length polymorphism (AFLP); the latter can also incorporate fluorescently labelled primers (FAFLP). Previous studies have confirmed that MRSA originated from the horizontal transfer of *SCCmec* elements into a limited number of meticillin-susceptible (MSSA) lineages, which has led to MRSA exhibiting a more clonal population structure compared to MSSA (Fitzgerald *et al.*, 2001; Musser & Kapur, 1992). Furthermore, Grundmann *et al.* (2010) utilized *spa* typing
to demonstrate that a limited number of such lineages have been selected for, and have subsequently spread into healthcare settings.

AFLP, a whole genome sampling methodology, detects polymorphisms within restriction endonuclease recognition-sequences in addition to insertions and deletions within core conserved and variable genomic regions (Vos et al., 1995). FAFLP enables precise fragment size detection and has thus improved reproducibility between laboratories. The application of FAFLP has enabled the discrimination of genetically related bacterial strains and has been successfully utilized in the analysis of outbreaks and epidemiological studies of a variety of bacterial genera (Desai et al., 1998; Grady et al., 1999, 2000; Janssen et al., 1996; Kober et al., 2011).

Comparative analysis studies of invasive and colonizing clinical isolates using AFLP or microarrays found no association between lineages and the severity of infection or single virulence or pathogenicity factors (Lindsay et al., 2006; Melles et al., 2004, 2009). AFLP data confirmed the clonal population structure of S. aureus that was revealed by MLST (Enright et al., 2000, 2002; Enright, 2003; Melles et al., 2004, 2009; Robinson & Enright, 2003) and highlighted the genetic heterogeneity within major lineages of S. aureus. Data showed that, whilst isolates of the same lineage shared a greater number of fragments compared to isolates from different lineages, unique AFLP markers could be identified for isolates within each specific cluster (Melles et al., 2009). With the advent of new technologies for the next-generation sequencing of whole genomes, it will be possible to detect these variations easily. However, it will be a number of years before this technique is used routinely, which, due to the cost and lack of technology, makes next-generation sequencing an unviable option in resource-poor settings.

This report presents a previously unpublished FAFLP methodology using a novel endonuclease combination to investigate the genetic diversity within MRSA. Using this methodology, amplified fragments (AFs) of a precise size, unique to isolates from successful lineages, were identified. These fragments were sequenced and the origin of the fragments and the features they encoded were identified. This will provide an insight into the predominance of particular lineages and enable the development of rapid and high-throughput assays for identifying MRSA lineages.

**METHODS**

**Bacteria, culture conditions and DNA extraction.** Seventy-one human MRSA isolates from the UK including 40 geographically diverse isolates referred to the England and Wales Staphylococcus reference laboratory (Health Protection Agency, UK) and 31 isolates from a hospital in Glasgow were studied. Six of the 71 isolates were obtained from a single patient with recurrent MRSA bacteraemia (over 6 months), seven isolates were from surveillance patients (four different centres) and 10 isolates were from patients with clinical symptoms (six separate centres). Of the remaining 48 isolates, eight were genome-sequenced strains and three were strains representative of each of the epidemic MRSA (EMRSA)-1, 2 and 3; no epidemiological data were available for the remaining 37 isolates.

All isolates were cultured on Columbia agar plates supplemented with horse’s blood (Oxoid) and incubated aerobically at 37 °C for 24 h. Genomic DNA was extracted using a Magna Pure LC Robot and MagNA Pure LC DNA Isolation kit III (Roche) according to the manufacturers’ recommendations and stored at –20 °C.

**SCCmec typing.** A previously described multiplex PCR assay was used to detect SCCmec types I–V (Milheirico et al., 2007). The presence of SCCmec type VI was detected using a protocol adapted from a real-time PCR assay described by Chen et al. (2009). A block-based multiplex PCR assay was developed and amplicons were resolved on a 2 % agarose gel. The amplicon sizes were used to infer the SCCmec type as described previously (Chen et al., 2009; Milheirico et al., 2007).

**MLST.** Standard MLST of S. aureus was performed as described previously (Enright et al., 2000). Sequencing reactions were performed using the BigDye Terminator v1.1 Cycle Sequencing kit and separated using an ABI 3730 automated capillary DNA sequencer (Life Technologies). The sequence of each amplicon was assigned to an allele by comparison to the MLST database (http://www.mlst.net) via BioNumerics 6.1 software (Applied Maths) (Platt et al., 2006). Novel alleles were submitted to the MLST database for confirmation.

**spa typing.** spa typing was performed by sequencing the polymorphic X region of protein A as described previously (Shopsin et al., 1999). Sequences were separated using an ABI 3730 automated capillary DNA sequencer and assigned to spa types using the Ridom SpaServer database (http://spa.ridom.de) via BioNumerics. Novel spa types were submitted to the Ridom SpaServer database.

**FAFLP.** FAFLP was performed empirically on a subset of isolates using three endonuclease combinations: EcoRI with Msel, HindIII with Hhal, and BglII with Csp61. The latter combination yielded amplified fragments evenly distributed in the size range 50–600 bp. A one-base selective BglII primer (BglII+X), labelled at the 5′-end with 5′-carboxyfluorescein (FAM), and an unlabelled non-selective Csp61 primer (Csp61+0) were used. The primer pair BglII+4 and Csp61+0 exhibited a higher discriminatory power (D) and hence this primer combination was used for FAFLP analysis of all isolates.

Genomic DNA (500 ng) was digested at 37 °C for 2 h in a final reaction volume of 20 µl consisting of 10 U Csp61, 2 µl 10 × Y+I/Tango buffer (Fermentas) and 0.3 µM DNAse-free RNase A (30 mg ml−1, Sigma-Aldrich). Three units of BglII (Fermentas) were subsequently added to each reaction mixture and incubated for a further 2 h at 37 °C. Ligation was performed by adding 27 µl solution containing 5 µl 10 × T4 DNA ligase buffer, 80 U T4 DNA ligase (New England Biolabs) and 0.8 µM (each) double-stranded adaptor (MWG Eurofins) to the double-digested DNA. The reaction was incubated at 12 °C for 16–18 h, heated to 65 °C for 10 min to inactivate the ligase and then cooled to 4 °C. The two strands of the Csp61 adaptor sequences were: 5′-TGTATCCTGAGAGCTCCTG-3′ and 5′-CATGACCTCTCGAGAAGCTTGA-3′ and 5′-GGACTAGTGGCTAGCCTATC-3′ and 5′-CTGATCTGAT-CGAGCAGTAGT5′. The reproducibility of the assay was evaluated on a subset of isolates by performing FAFLP analysis in
Fragment analysis. FAFLP fragments were separated on an ABI 3730 sequencer by loading a denatured mixture of 1 µl PCR product, 10 µl Hi-Di formamide and 0.5 µl LIZ600 internal lane marker (Life Technologies). Fluorescent AFs were sized using GeneMapper 4.0 software (Life Technologies). The presence or absence of fragments was recorded in a binary format and imported into BioNumerics. The Jaccard co-efficient of similarity was calculated with a band position tolerance of 0.6 % and cluster analysis was performed by the UPGMA method. Error flags and cophenetic correlations, which inferred the stability and confidence of each cluster, respectively, were calculated in BioNumerics. In-silico FAFLP profiles were generated for genome-sequenced strains using AFLIE, an AFLP fragment predictor program (Underwood, 2011). The percentage concordance between in-silico and experimental profiles for the eight sequenced strains was determined.

Identification of lineage-specific genetic variation. Unique fragments of precise sizes (lineage-specific AFs) were identified in FAFLP profiles of isolates belonging to one of six clonal lineages (CC1, 5, 8, 22, 30 or 45). Lineage-specific AFs were amplified from these six lineages using the selective primers BglII + A and Csp61 + A. Fragments were resolved on a 4 % agarose gel, purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and sequenced as mentioned earlier in the MLST methods section. Sequences of the AFs were analysed in BioEdit 7.0.9 software and BLAST searches were performed on the NCBI database. Fragment-specific primers were designed from the sequence data using the Primer3 program (Rozen & Skaltsky, 2000). Individual PCRs were designed for each fragment to determine the lineage specificity of the fragments (Table S1, available in JMM Online). PCRs were performed on isolates from each of the clonal lineages CC1, 5, 8, 22, 30, 45, 59 and 97. Amplicons obtained for each fragment from isolates of the eight clonal complexes (CCs) were sequenced to determine the sequence variation within each lineage and sequences were analysed in BioEdit.

Discriminatory power. The discriminatory power (D) based on Simpson’s index of diversity (Hunter & Gaston, 1988; Simpson, 1949) and the 95 % confidence interval (CI) as proposed by Grundmann et al. (2001) were calculated for MLST, SCCmec, spa and FAFLP data.

RESULTS

FAFLP analysis
The primer combination BglII + A and Csp61 + 0 yielded 46–68 AFs (size calling tolerance ± 0.5 bp) ranging in size from 50 to 600 bp per isolate. This combination of AFs constituted a FAFLP profile and covered ~0.56 % of the genome. Reproducibility results exhibited ≥98 % similarity (two or less fragment difference) between the duplicate profiles generated for each of the isolates examined. Based on these results, FAFLP profiles were considered to be identical when ≤3 AF differences were exhibited. Sixty-seven FAFLP profiles were exhibited amongst the 71 isolates; a total of 304 distinct AFs were identified and 295 of these were polymorphic.

Cluster analysis
A dendrogram was generated from BglII + A and Csp61 + 0 FAFLP data (Fig. 1). Based on MLST data, which assigned the 71 isolates to 12 CCs and one singleton, a 70 % similarity cut-off was applied to the FAFLP dendrogram. Twelve FAFLP clusters were identified among 69 MRSA isolates and the two remaining CC88 isolates exhibited divergent profiles (>30 % difference).

Cluster 1 consisted of 18 isolates assigned to CC22. FAFLP differentiated isolates within this cluster, exhibiting 2–22 % divergence and up to 85 polymorphic AFs. Cluster 6 contained five isolates exhibiting up to 29 % divergence and up to 65 polymorphic fragments. Four of these isolates belonged to CC1 and one to CC7 (novel single locus variant of ST7). Eleven isolates of CC5 (cluster 7) displayed up to 98 polymorphic AFs which constituted an 8–22 % difference. Clusters 3 and 4 consisted of eight and six CC8 isolates exhibiting up to 97 and 58 AF differences, respectively. The majority of the isolates within cluster 3 (88 %) belonged to ST8 or ST250, whilst all isolates within cluster 4 belonged to ST239. Up to 66 % similarity (17 common AFs) was exhibited between isolates of the two clusters, whilst up to 28 % difference was observed within each of the clusters.

Of the five CC30 isolates in cluster 9, four belonged to ST36 and one to ST30. Up to 58 polymorphic AFs were identified, constituting a ≤14 % difference. The single ST30 isolate diverged from the other isolates in the cluster by 10 % and up to 33 polymorphic AFs. The five isolates within CC45 (cluster 11) showed a ≤26 % difference attributed to up to 77 polymorphic AFs. Isolates within clusters 2, 5, 8, 10 and 12 consisted of ≤3 isolates and exhibited <20 % difference. The isolates from each of these clusters belonged to CC672, 80, 97, 59 and ST93, respectively.

A 92 % similarity cut-off was further applied to the dendrogram to reflect the FAFLP-based divergence between the isolates. Within the CC22 isolates (cluster 1), four FAFLP ‘subtypes’ were identified. Within each subtype, 2–43 AF differences were observed; conversely 4–16 AF differences were observed between subtypes. Within CC30 isolates (cluster 9), two subtypes were identified and one subtype comprised all ST36 isolates.

Of the six bacteraemia isolates, four were CC22 isolates (cluster 1) and two belonged to CC361 (cluster 2). Isolates of the same CC were closely related to one another and exhibited identical spa and SCCmec types. The seven surveillance isolates were assigned to CC5, 22 and 80 and were found in clusters 3, 1 and 5 respectively. Two isolates from the same centre clustered together. FAFLP assigned nine of the 10 clinical isolates to five different clusters, whilst one did not belong to any cluster; these isolates were assigned to seven CCs. Surveillance and clinical isolates belonged to similar genetic backgrounds but did not cluster according to the epidemiological data. The eight genome-sequenced strains clustered according to their CC assignment. Comparisons of the FAFLP profiles of these strains with in-silico profiles revealed 75–82 % concordance.
Fig. 1. Dendrogram showing the genetic relationship between FAFLP profiles and clonal complexes. The dendrogram was derived from FAFLP data using the primer pair BglII+ and Csp6I+ with UPGMA cluster analysis. FAFLP clusters were designated arbitrary cluster numbers 1–12 based on a 70 % sequence similarity cut-off indicated by the black line on the left. The black line to the right represents a 92 % sequence similarity cut-off, exhibiting FAFLP ‘subtypes’ within the clusters. The clonal complex assignment of isolates within each cluster is shown preceding each corresponding node. Horizontal grey bars indicate error flags representing the SD for each cluster. Values at nodes represent cophenetic correlations, which compare the dendrogram similarities with the similarity matrix values. The spa type and SCCmec type of each isolate are shown on the right. An asterisk indicates that the isolate was positive only for mecA, mecI, ccr3 and ccrC.
The index of discriminatory power (D) was 0.998 for FAFLP with a CI of 0.996–1.000. This was higher compared to that of spa typing, which was 0.975 with a CI of 0.964–0.986, MLST, which was 0.909 with a CI of 0.868–0.949, and SCCmec typing, which was 0.71 with a CI of 0.604–0.795. None of the calculated CI values overlapped between the four typing techniques. A limited number of MLST sequence types (1–4) and SCCmec variants (2–5) were detected within each CC. Taking the number of isolates available from each CC into consideration, the largest number of spa genotypes was demonstrated within CC5, CC30 and CC45. The highest numbers of FAFLP profiles were exhibited within CC5, 8, 22 and 45. In both cases, FAFLP differentiated within isolates belonging to a single spa type (Fig. 1).

Identification of lineage-specific variation

FAFLP analysis using BgII+A and Csp6I+A primers identified 25 AFs of different sizes that were specific to isolates belonging to one of five major lineages (CC5, 8, 22, 30 or 45); no AF specific to CC1 was identified. The greatest number of specific AFs was observed within isolates of CC30 followed by CC22 and CC5. Fragment-specific PCRs revealed the presence of similar fragment sequences in isolates of CC1, 59 and 97, in addition to the five lineages mentioned above. Thirteen of 25 AFs were sequenced from isolates representative of the eight CCs; the remaining 12 AFs were mixed with AFs of a similar size. Five of the 13 sequences were conserved and displayed minor variations between all isolates. Within the remaining eight sequences (F1–F8), a combination of single nucleotide polymorphisms (SNPs), point mutations and indels (hereafter referred to as SNPs) were displayed. The majority of SNPs within these sequences were conserved between isolates of a single CC (Table 1). The eight sequences exhibited a combination of unique SNPs within the eight CCs. The number of variations specific to each CC was also variable amongst each sequence. The highest number of specific SNPs was observed in CC45 isolates (68 SNPs) whilst the lowest number of SNPs was observed within isolates of CC8 (3 SNPs) (Table 1). Within each of the sequences, the variations were specific to a different number of CCs and types of CC. Again, the CC30 (F4)-specific AF harboured variations within the highest number of CCs (8) whilst the lowest number (1) was observed within the CC45 (F8)-specific AF.

BLAST analysis of the fragments F2, F4, F5 and F7 indicated that these regions partially encode functions relating to the repair, replication and recombination of the genome. The analysis of fragment F1 revealed that this region originates from the SCCmec insertion site and shows sequence similarity to the region encoding the enzyme transposase. Fragments F3, F6 and F8 encoded, in part, factors that can induce the cell to undergo apoptosis, whilst F4 and F5 partially encoded acetolactate synthase, an enzyme essential for the biosynthesis of certain amino acids (Table 1).

DISCUSSION

The genetic diversity between and within 13 MRSA lineages was investigated using FAFLP analysis. Cluster analysis of FAFLP data differentiated 12 of the 13 MRSA lineages and the clusters were concordant with the MLST CCs. Size-specific AFs were identified in isolates of five predominant MRSA lineages but sequence data indicated the presence of the AFs in all eight CCs examined. However, lineage-specific sequence variations in the form of SNPs, point mutations and indels were identified within the eight sequenced fragments. A combination of SNPs found within the fragment sequences constituted the lineage specificity. Interestingly, the largest number of variations (68 SNPs) was observed within CC45 isolates as compared to three SNPs in CC8 isolates; CC45 is a lineage that is less predominant than CC22 or CC30 in the UK, whereas CC8 has been predominant in the UK. This probably indicates that CC45 isolates are under selective pressure compared to CC8 isolates, which are relatively stable at the eight loci sampled in the genome by FAFLP. However, other factors such as changes in the population dynamics and clonal expansion could also account for this diversity. Within the eight sequenced fragments, the number of sequence variations between each of the fragments, and also between each CC, varied. The distribution of lineage-specific sequence variations between and within CCs indicates both the genetic diversity around the genome within isolates of the same lineage and the diversity between lineages in the same genomic region. Insights such as these can aid in the surveillance of emerging epidemics.

The identified AFs partially encoded toxins, antibiotic resistance determinants, bacteriophage resistance factors and a transposase within the SCCmec insertion site region; the latter harbours numerous virulence and antibiotic resistance determinants (Malachowa & DeLeo, 2010; Noto et al., 2008). Of the eight AFs sequenced, two encoded, in part, a recombination protein, RecR, and a DNA photolyase, both of which are essential for the repair of damaged DNA (Inoue et al., 2008; Sancar, 2003). A further two AFs partially encoded the enzyme acetolactate synthase; mutations within this enzyme result in the inability of the bacterium to grow in certain conditions (Dailey & Cronan, 1986). The sequences showed a high degree of variation, indicating that this region is under selective pressure in numerous lineages. An abortive infection bacteriophage resistance protein was also partially encoded on two AFs. Mutations within this region result in thickening of the cell wall, thus preventing the penetration of antibiotic molecules and bacteriophages and often results in programmed cell death (Frankel et al., 2010; Garvey et al., 1995). As this region may play a role in pathogenicity, the CC8-specific SNP identified in this sequence may provide an insight into the predominance of
The remaining AF encoded, in part, an addiction module toxin/antitoxin that has a role in programmed cell death under conditions of nutritional stress or infection by bacteriophages (Engelberg-Kulka & Glaser, 1999). In genomic regions, such as these, that encode vital cell functions, we would expect minor genetic variations as these could result in significant phenotypic changes that may contribute to strain survival, fitness and success.

A previous study by Savelkoul et al. (2007) sequenced polymorphic AFLP fragments between MRSA and MSSA and one fragment showed sequence similarity to a transposase, similar to the fragment F1, identified in our study. Our results are concordant with the findings of Melles et al. (2004) which suggested that AFLP markers become lineage-specific via the accumulation of point mutations rather than larger genetic events such as transposition. These point mutations are found within fragments that encode proteins related to virulence and pathogenicity. The lineage-specific variations may, therefore, account for changes in the virulence and pathogenicity between major CCs.

In general, the FAFLP data were found to be concordant with MLST sequence types. Identical and polymorphic AFs between isolates can indicate regions of genetic similarity or variation and the level of genetic relatedness between MLST sequence types. The FAFLP profiles of CC8 isolates (clusters 3 and 4, Fig. 1) shared the most number of AFs between different clusters. FAFLP split the ST239 isolates (cluster 4, Fig. 1) from the majority of the other CC8 sequence types. A previous study by Robinson & Enright (2004) indicated a significant genetic transfer event between an ST8 and ST30 progenitor to form ST239. FAFLP identified 17 AFs (25–37% of each profile) common to all isolates of CC8, indicating the origin of these fragments as regions of genetic similarity between ST239 and the remaining CC8 sequence types.

This report demonstrates FAFLP to be a reproducible and rapid technique. FAFLP exhibits a higher discriminatory power compared to techniques such as spa typing, although both techniques such as spa typing samples variation at a single locus within the genome, whereas FAFLP samples the whole genome. It should be noted, however, that FAFLP data may vary depending on the sequencing analyser used and is less portable as it is based on fragment data and therefore is more vulnerable to data interpretation than sequence data.

### Table 1. Variation and function of lineage-specific fragments

<table>
<thead>
<tr>
<th>Lineage-specific fragments (CC)</th>
<th>Total no. SNPs per fragment</th>
<th>Number of SNPs*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC1</td>
<td>CC3</td>
</tr>
<tr>
<td>F1 (CC5)</td>
<td>16</td>
<td>3†</td>
</tr>
<tr>
<td>F2 (CC22)</td>
<td>34</td>
<td>–</td>
</tr>
<tr>
<td>F3 (CC30)</td>
<td>27</td>
<td>–</td>
</tr>
<tr>
<td>F4 (CC30)</td>
<td>70</td>
<td>8‡</td>
</tr>
<tr>
<td>F5 (CC30)</td>
<td>28</td>
<td>–</td>
</tr>
<tr>
<td>F6 (CC30)</td>
<td>16</td>
<td>1†</td>
</tr>
<tr>
<td>F7 (CC45)</td>
<td>18</td>
<td>–</td>
</tr>
<tr>
<td>F8 (CC45)</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Total number of SNPs per CC</td>
<td>–</td>
<td>12</td>
</tr>
</tbody>
</table>

*SNPs, includes number of lineage-specific single nucleotide polymorphisms, point mutations and indels.

†SNPs identified are ST1 specific only.

‡Four of eight SNPs identified are ST1 specific only.

In general, the FAFLP data were found to be concordant with MLST sequence types. Identical and polymorphic AFs between isolates can indicate regions of genetic similarity or variation and the level of genetic relatedness between MLST sequence types. The FAFLP profiles of CC8 isolates (clusters 3 and 4, Fig. 1) shared the most number of AFs between different clusters. FAFLP split the ST239 isolates (cluster 4, Fig. 1) from the majority of the other CC8 sequence types. A previous study by Robinson & Enright (2004) indicated a significant genetic transfer event between an ST8 and ST30 progenitor to form ST239. FAFLP identified 17 AFs (25–37% of each profile) common to all isolates of CC8, indicating the origin of these fragments as regions of genetic similarity between ST239 and the remaining CC8 sequence types.

Acetolactate synthase

DNA photolyase family

Abortive infection bacteriophage resistance protein

Transposase in SCCmec insertion site region

Recombination protein RecR

Abortive infection bacteriophage resistance protein

Acetolactate synthase

Addiction module toxin, Txe/YoeB family and antitoxin, Axe family

In general, this lineage as an epidemic MRSA strain in the UK. The genome regions, such as these, that encode vital cell functions, may contribute to strain survival, fitness and success.

Polymeric AF markers of MRSA and MSSA may contribute to strain survival, fitness and success. The results of this study are concordant with the findings of Melles et al. (2004) which suggested that AFLP markers become lineage-specific via the accumulation of point mutations rather than larger genetic events such as transposition. These point mutations are found within fragments that encode proteins related to virulence and pathogenicity. The lineage-specific variations may, therefore, account for changes in the virulence and pathogenicity between major CCs.

A previous study by Savelkoul et al. (2007) sequenced polymorphic AFLP fragments between MRSA and MSSA and one fragment showed sequence similarity to a transposase, similar to the fragment F1, identified in our study. Our results are concordant with the findings of Melles et al. (2004) which suggested that AFLP markers become lineage-specific via the accumulation of point mutations rather than larger genetic events such as transposition. These point mutations are found within fragments that encode proteins related to virulence and pathogenicity. The lineage-specific variations may, therefore, account for changes in the virulence and pathogenicity between major CCs.

In general, the FAFLP data were found to be concordant with MLST sequence types. Identical and polymorphic AFs between isolates can indicate regions of genetic similarity or variation and the level of genetic relatedness between MLST sequence types. The FAFLP profiles of CC8 isolates (clusters 3 and 4, Fig. 1) shared the most number of AFs between different clusters. FAFLP split the ST239 isolates (cluster 4, Fig. 1) from the majority of the other CC8 sequence types. A previous study by Robinson & Enright (2004) indicated a significant genetic transfer event between an ST8 and ST30 progenitor to form ST239. FAFLP identified 17 AFs (25–37% of each profile) common to all isolates of CC8, indicating the origin of these fragments as regions of genetic similarity between ST239 and the remaining CC8 sequence types.

Acetolactate synthase

DNA photolyase family

Abortive infection bacteriophage resistance protein

Transposase in SCCmec insertion site region

Recombination protein RecR

Abortive infection bacteriophage resistance protein

Acetolactate synthase

Addiction module toxin, Txe/YoeB family and antitoxin, Axe family

Our data suggest that FAFLP analyses conserved regions in a similar manner to MLST and can be utilized to infer the phylogeny of each MSA lineage. CC2 is a clonal lineage, with the exception of whole genome sequencing, MLST is the only effective technique to subtype this lineage. However, FAFLP proved to be a reproducible and rapid technique, with the exception of whole genome sequencing, can effectively subtype this lineage. In this study, we used four AFLP markers to probe the genetic heterogeneity within the lineage and identified four subtypes (cluster 1, Fig. 1). This report demonstrates FAFLP to be a reproducible and rapid technique. FAFLP exhibits a higher discriminatory power compared to techniques such as spa typing, although both techniques such as spa typing samples variation at a single locus within the genome, whereas FAFLP samples the whole genome. It should be noted, however, that FAFLP data may vary depending on the sequencing analyser used and is less portable as it is based on fragment data and therefore is more vulnerable to data interpretation than sequence data.

In general, the FAFLP data were found to be concordant with MLST sequence types. Identical and polymorphic AFs between isolates can indicate regions of genetic similarity or variation and the level of genetic relatedness between MLST sequence types. The FAFLP profiles of CC8 isolates (clusters 3 and 4, Fig. 1) shared the most number of AFs between different clusters. FAFLP split the ST239 isolates (cluster 4, Fig. 1) from the majority of the other CC8 sequence types. A previous study by Robinson & Enright (2004) indicated a significant genetic transfer event between an ST8 and ST30 progenitor to form ST239. FAFLP identified 17 AFs (25–37% of each profile) common to all isolates of CC8, indicating the origin of these fragments as regions of genetic similarity between ST239 and the remaining CC8 sequence types.
macro-evolution of MRSA lineages (Lindsay & Holden, 2004). Moreover, whilst FAFLP probably underestimates the level of genetic diversity in accessory regions, it does sample such variable regions and can help facilitate our understanding of the micro-evolution of S. aureus (Melles et al., 2004; van Leeuwen et al., 2005). Existing molecular typing techniques for MRSA may not detect minor variations throughout the genome. Although next-generation sequencing provides the ultimate resolution for detecting genetic diversity, it is not yet feasible for the majority of laboratories to resource the techniques and carry out the necessary, often problematic, analysis. As such, it will be a number of years before next-generation sequencing is utilized routinely.

The comparison of profiles between isolates may also provide insight into the population genetics of MRSA and the evolution of successful lineages. FAFLP has been demonstrated as a useful tool to investigate the genetic diversity between MRSA CCs. The lineage-specific markers will be utilized in the development of diagnostic PCR-based assays for MRSA typing. Such assays would prove to be invaluable high-throughput and rapid tools that would be economical for the majority of resource-poor laboratories. Further work is needed to investigate the genetic diversity within MSSA isolates and to identify whether similar lineage-specific fragment sizes and variations are present.

ACKNOWLEDGEMENTS

A proportion of the data was previously presented at the 20th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) on 12th April 2010 in Vienna, Austria, and the 21st ECCMID on 9th May 2011 in Milan, Italy. This work was funded by a PhD studentship from the Health Protection Agency.

REFERENCES


