Distribution and characterization of staphylococcal interspersed repeat units (SIRUs) and potential use for strain differentiation

K. J. Hardy,1,2 D. W. Ussery,3 B. A. Oppenheim1 and P. M. Hawkey1,2

1West Midlands Public Health Laboratory, Health Protection Agency, Heartlands Hospital, Bordesley Green East, Birmingham B9 5SS, UK
2Division of Infection and Immunity, University of Birmingham, Birmingham B15 2TT, UK
3Center for Biological Sequence Analysis, Biocentrum-DTU, Building 208, The Technical University of Denmark, DK-2800 Denmark

Variable-number tandem repeats (VNTRs) have been shown to be a powerful tool in the determination of evolutionary relationships and population genetics of bacteria. The sequencing of a number of Staphylococcus aureus genomes has allowed the identification of novel VNTR sequences in S. aureus, which are similar to those used in the study of the evolution of Mycobacterium tuberculosis clades. Seven VNTRs, termed staphylococcal interspersed repeat units (SIRUs), distributed around the genome are described, occurring in both unique and multiple sites, and varying in length from 48 to 159 bp. Variations in copy numbers were observed in all loci, within both the sequenced genomes and the UK epidemic methicillin-resistant S. aureus (EMRSA) isolates. Clonally related UK EMRSA isolates were clustered using SIRUs, which provided a greater degree of discrimination than multi-locus sequence typing, indicating that VNTRs may be a more appropriate evolutionary marker for studying transmission events and the geographical spread of S. aureus clades.

INTRODUCTION

A number of different tandem repeat sequences have been described in both prokaryotes and eukaryotes. All have a unique length and DNA sequence of the repeat unit, different copy numbers per genome, and different organization within the genome. Loci with repeat units of 6–100 bp are referred to as minisatellites, whilst loci with smaller repeat units are termed microsatellites (Jeffreys et al., 1992; Tautz & Schlotterer, 1994). The tandem repeat nature of minisatellites has been studied extensively in eukaryotic genomes. The first bacterial species in which they were identified was Mycobacterium tuberculosis, being described as mycobacterial interspersed repeat units (MIRUs) (Supply et al., 1997, 2000). These loci have a characteristic structure consisting of a variable number of near-identical repeated DNA sequences arranged consecutively. A general term for these repeats is variable-number tandem repeats (VNTRs), and true VNTRs have now been reported in these repeats is variable-number tandem repeats (VNTRs), DNA sequences arranged consecutively. A general term for consisting of a variable number of near-identical repeated units (SIRUs) and potential use for strain differentiation

INTRODUCTION

A number of different tandem repeat sequences have been described in both prokaryotes and eukaryotes. All have a unique length and DNA sequence of the repeat unit, different copy numbers per genome, and different organization within the genome. Loci with repeat units of 6–100 bp are referred to as minisatellites, whilst loci with smaller repeat units are termed microsatellites (Jeffreys et al., 1992; Tautz & Schlotterer, 1994). The tandem repeat nature of minisatellites has been studied extensively in eukaryotic genomes. The first bacterial species in which they were identified was Mycobacterium tuberculosis, being described as mycobacterial interspersed repeat units (MIRUs) (Supply et al., 1997, 2000). These loci have a characteristic structure consisting of a variable number of near-identical repeated DNA sequences arranged consecutively. A general term for these repeats is variable-number tandem repeats (VNTRs), and true VNTRs have now been reported in Bacillus anthracis, Legionella pneumophila, Pseudomonas aeruginosa, Salmonella enterica and Escherichia coli O157 (Keim et al., 2000; Liu et al., 2003; Noller et al., 2003; Onteniente et al., 2003; Pourcel et al., 2003). With the exception of M. tuberculosis, these studies have been confined to comparisons of molecular strain differentiation techniques. Staphylococcus aureus, and in particular methicillin-resistant S. aureus (MRSA), is an important nosocomial pathogen. In order to understand the molecular evolution of different strains in relation to important physiological properties such as pathogenicity and survival, evolutionary markers are needed. Multi-locus sequence typing (MLST) examines the slowly evolving genomic core, and is a useful tool for defining chromosomal ancestry (Enright et al., 2000). Through use of this technique and others, the population structure of S. aureus has been studied, and it has been shown to consist of a limited number of clones in comparison to other bacterial species. Investigations into the evolution of MRSA have shown that horizontal transfer of the SCCmec complex, containing the mecA gene, has occurred on numerous occasions (Enright et al., 2002). Despite this, the genetic backgrounds of hospital MRSA isolates are narrow in comparison to community methicillin-sensitive S. aureus (MSSA) and the recently described community MRSA (Enright et al., 2002). Therefore, more rapidly evolving genetic markers are needed to study hospital MRSA.
The sequences of seven *S. aureus* genomes are now available; the sequences of N315, MW2, Mu50, MSSA 476 and MRSA 252 are published (Baba *et al.*, 2002; Kuroda *et al.*, 2001; Holden *et al.*, 2004), whilst the sequences of NCTC 8325 and COL are currently being annotated (http://www.genome.ou.edu/staph.html and http://www.tigr.org/tdb/mdb/mdbinprogress.html respectively). Five of the strains (COL, MRSA 252, Mu50, N315 and MW2) are MRSA, whilst MSSA 476 and NCTC 8325 are mecillinam sensitive. N315 and Mu50 are closely related strains that are hospital-acquired and vancomycin-intermediate-resistant respectively, whilst MRSA 252 is an epidemic hospital strain, predominant within the UK. MW2 is a hypervirulent community strain isolated in the USA, but it has been associated with MSSA 476, a community isolate from the UK. The sequencing of these genomes has allowed us to identify novel multiple tandem repeats that are similar to MIRUs in *M. tuberculosis*, and the following report describes the nature and distribution of these repeats, and their potential as a novel tool for understanding the micro-evolution of hospital MRSA.

**METHODS**

**Tandem repeat locus identification.** Three of the sequenced *S. aureus* genomes, MW2, N315 and MRSA 252, were searched for the presence of tandem repeats using the tandem repeat finder program (http://c3.biomath.mssm.edu/trf.html).

The specified search criteria were that there must be > 94 % conservation between the repeat units, and that the repeat unit must be between 45 and 200 bp in length. A total of 26 tandem repeats across the three genomes fitted these criteria, with seven identified in EMRSA 16, 13 in MW2 and six in N315. The presence of these repeat units, and the degree of similarity of the flanking regions in the other sequenced *S. aureus* genomes, were investigated using BLAST. Seven of the repeat units were present in all seven sequenced genomes, and they had a high degree of similarity in the flanking regions; these were designated staphylococcal interspersed repeat units (SIRUs) (Table 1). The presence of SIRU01, 05, 07, 13, 15 and 21 was studied in a collection of the most widely distributed nosocomial epidemic MRSA (EMRSA) strains in the UK.

**Primer design.** Primers were designed to anneal within the conserved flanking regions outside of the repeat regions for all seven SIRUs. The length of flanking region included within the amplicon ranged from 103 to 260 bp. Primers were designed using the computer program Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3 www.cgi). In order to be able to amplify all DNA templates using the same reaction conditions, primer length and G + C content were kept as constant as possible. A common magnesium concentration and annealing temperature were identified using serial titrations.

**Isolates and DNA preparation.** A collection of the UK EMRSA types (1–16; NCTC), all of which have been previously described and typed using PFGE (Kerr *et al.*, 1990; Moore & Lindsay, 2002), were used to assess the discriminatory ability of the repeat units. DNA was extracted as described by Kumari *et al.* (1997).

**PCR amplification and detection.** Briefly, each PCR reaction (25 μl) contained 1× PCR Gold Buffer (Applied Biosystems), 200 μM each deoxynucleoside triphosphate (dNTP), 1 μM each flanking primer, 2 mM MgCl2, 0·5 U AmpliTaq Gold (Applied Biosystems) and 2 μl DNA. Amplification was performed in a Multi-Block System (MBS) thermal cycler (ThermoHybaid). An initial denaturation temperature of 96 °C for 2 min was followed by 30 cycles of denaturation at 96 °C for 30 s, annealing at 55 °C, and elongation at 72 °C for 30 s. There was a final elongation step at 72 °C for 2 min. The amplicons were analysed on a 2 % (w/v) agarose gel (Invitrogen) for 2·5 h at 150 V, with 50 bp ladder size standards. Gels were stained using ethidium bromide, visualized with UV and photographed. The size of each amplicon was determined, and the number of repeats calculated, taking into account the size of the repeat unit and the flanking region.

**DNA sequencing and PFGE.** The DNA sequences of all seven SIRUs from EMRSA-16 were determined using Big Dye PCR reaction (Applied Biosystems), and analysed on a Primus HT 3700 DNA analyser by the functional Genomics Laboratory, University of Birmingham, UK. PFGE was carried out by using a standardized method with Smal restriction endonuclease (Invitrogen) as described by Murchan *et al.* (2003).

**Cluster analysis.** SIRU and PFGE results were entered into BioNumerics (Applied Maths), analysed using Euclidian coefficient, and displayed via the unweighted pair group method using arithmetic averages (UPGMA).

**RESULTS**

We investigated seven sequenced *S. aureus* genomes for the presence of VNTRs. Seven VNTRs, termed SIRUs, were identified. Each SIRU was made up of repeated sequences, the number of which varied from strain to strain. The seven SIRUs were found to be distributed around the genome of *S. aureus* (Table 1).

### Table 1. Genomic characteristics of SIRUs identified from sequenced *S. aureus* genomes

<table>
<thead>
<tr>
<th>Locus no.</th>
<th>Genome tandem repeat finder performed on</th>
<th>Repeat length (bp)</th>
<th>Conservation (%)</th>
<th>G + C content (mol%)</th>
<th>Location on sequenced genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>EMRSA 16</td>
<td>55</td>
<td>98</td>
<td>40</td>
<td>384017–384123</td>
</tr>
<tr>
<td>05</td>
<td>EMRSA 16</td>
<td>60</td>
<td>98</td>
<td>36</td>
<td>2488212–2488383</td>
</tr>
<tr>
<td>07</td>
<td>EMRSA 16</td>
<td>56</td>
<td>95</td>
<td>46</td>
<td>2855668–2855783</td>
</tr>
<tr>
<td>13</td>
<td>MW2</td>
<td>64</td>
<td>100</td>
<td>30</td>
<td>1215469–1215625</td>
</tr>
<tr>
<td>15</td>
<td>MW2</td>
<td>131</td>
<td>99</td>
<td>32</td>
<td>1725259–1725542</td>
</tr>
<tr>
<td>16</td>
<td>MW2</td>
<td>159</td>
<td>98</td>
<td>35</td>
<td>1825755–1826239</td>
</tr>
<tr>
<td>21</td>
<td>N315</td>
<td>48</td>
<td>94</td>
<td>45</td>
<td>122940–123176</td>
</tr>
</tbody>
</table>
Occurrence of repetitive sequences in sequenced genomes

The presence and position of the seven SIRUs were investigated in each of the seven sequenced S. aureus genomes. Interestingly two different phenomena were observed, in that four of the SIRUs were present only once on the genome (single-site SIRUs), whilst three SIRUs were present several times around the genome (multiple-site SIRUs) (Fig. 1). The positions of SIRUs 13, 15, 16 and 21, which occurred only once, were different relative to each other around the genome (Fig. 1). The positioning of single-site SIRUs remained constant between different genomes.

These four single-site SIRUs had a very high percentage similarity in both the 3' and 5' flanking sequences. The multiple-site SIRUs 01 and 07 occurred at multiple sites around the genome. Variation in the number of the repeat units present at each site on the genome was observed, as well as some variation in the repeat sequence between sites. However, the flanking regions of each of the multiple-site SIRUs present on the same genome had very little similarity to each other. Similarity in flanking regions for specific sites was observed from genome to genome. The multiple-site SIRU that was selected for analysis contained the repeat units with the highest sequence homology, and all strains had flanking regions with high similarity. SIRU05 was present in two places on the genome. Both MW2 and MSSA 476 contained the repeat unit, and had a flanking region with 100% similarity at the 3' end, but the flanking region at the 5' end was significantly different.

Comparison of numbers of repeat units and DNA sequences of SIRUs between different sequenced genomes

Length polymorphisms were seen at all loci (Table 2). The number of repeat units present in a genome was calculated on full repeat units only. SIRU01, 05, 15 and 16 all had full repeats (see Fig. 2 for SIRU05). When partial repeats were present, they were included in the flanking region for determination of the number of repeat units. For example, in SIRU15, an extra 22 bp of the repeat unit was present in all genomes, but in the calculation this was included in the flanking region (Fig. 3). A more complicated situation was observed in SIRU13 (Fig. 4), which had half-repeats present in all genomes. COL and NCTC 8325 had only half a repeat unit, and were therefore classified as having no full repeats in the calculations (Table 1).

There was a high level of sequence identity between the different sequenced genomes within the repeat units, with a few exceptions. The highest degrees of sequence identity in all repeat units were seen in SIRU05, 13 and 21. In contrast, COL, NCTC 8325 and MRSA 252 had a repeat unit

![Genome atlas of MRSA strain N315. The genomic position of each of the SIRUs is represented on the grey bands by red (positive DNA strand) and blue (negative DNA strands) markers. Between SIRU07 and 13, the genes on the forward and reverse strands are shown in red and blue respectively. In the centre circle, the number of bases around the genome is shown; 0M, the origin of replication.](http://mic.sgmjournals.org)

**Table 2.** Complete repeat units in the seven S. aureus sequenced genomes as shown by sequence alignment

<table>
<thead>
<tr>
<th>Locus</th>
<th>NCTC 8325</th>
<th>COL</th>
<th>MW2</th>
<th>Mu50</th>
<th>N315</th>
<th>MRSA 252</th>
<th>MSSA 476*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

 Allelic profile: 3630145 4620145 1532234 2313135 2313135 2321225 15–1214

*–, The flanking region and the repeat unit were not present.
that was identical in length in SIRU15, but differed by 26% from the consensus sequence (Fig. 3). N315 and Mu50 have been shown to be clonal (Kuroda et al., 2001), and investigations of their SIRUs show that they have the same number of repeat units in all loci. They also contain the same base changes in all repeat units. Similarity in the base changes observed in the repeat units has been seen in other SIRU repeat units, for example in SIRU16 of NCTC 8325 and COL.

The repeat units were true tandem repeats in all SIRUs, with no intervening sequence between the repeats, apart from SIRU01. SIRU01 had either 1 or 2 bp between the first and second repeats in all genomes. When sizing a fragment using a gel-based method, the presence of 1 or 2 bp does not change the apparent number of copies present.

**Position of repeat**

The published *S. aureus* sequences have been annotated, and we were able to establish the position of the SIRUs.
in relation to coding and non-coding regions. Five of the seven SIRUs were in non-coding regions. SIRU13 was in the coding region for MW1681, a hypothetical protein. SIRU21 was in the coding region for protein A, spa. This region has been previously used for epidemiological typing (Koreen et al., 2004). Variation in the number of spa repeats is based on a repeat size of 24 bp; our software identified a repeat of 48 bp. SIRU15 and 16 were in non-coding regions, but were close to the obg and hmrB genes respectively.

Use of repeats in defining the genetic relatedness of EMRSA types

PCR was performed on the EMRSA isolates from the UK which have been previously shown to differ by varying degrees by PFGE and MLST, to evaluate the potential discriminatory ability of the loci. Of the 16 isolates typed, 12 different SIRU profiles were observed, with different degrees of variation within the SIRU profile. Length polymorphisms of the amplified products were seen in all loci, with the greatest variations being in SIRU05, 13 and 21 (Table 3). On amplification, a single band was obtained for all reactions, confirming the specificity of the primers, and reproducibility was demonstrated, with a product of the same length being obtained on repetitive amplification of the same isolates. The sequence data obtained from sequencing all seven SIRUs from EMRSA16 and comparing them to the published EMRSA16 genome (MRSA 252) showed 100% concordance. Further confirmation of the correct amplification of each of the SIRUs was demonstrated by the fact that all our amplicon sizes fell within the predicted size for an exact number of repeats.

SIRU typing had the ability to discriminate between the EMRSA strains (Fig. 5). A cluster of ten EMRSA isolates containing five different SIRU profiles differed by a maximum of two loci, and were clustered into three sequence types (STs) by MLST. The STs contained within this cluster only differed at one locus, either pta (phosphate acetyltransferase) or arcC (carbamate kinase). For this set of ten isolates, the MLST types and SIRU profiles were not always in concordance. The remaining six EMRSA isolates had distinct SIRU profiles and MLST types, and a greater degree of variation within the loci was observed.

**DISCUSSION**

The value of VNTRs as a powerful tool in the investigation of both population genetics and evolutionary events has

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Locus no.*</th>
<th>Allelic profile*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>01 13 15 16 21 05 07</td>
<td></td>
</tr>
<tr>
<td>EMRSA 1</td>
<td>2 1 1 3 3 2 2</td>
<td>2113322</td>
</tr>
<tr>
<td>EMRSA 2</td>
<td>2 1 1 3 1 3 1</td>
<td>2113313</td>
</tr>
<tr>
<td>EMRSA 3</td>
<td>1 3 1 3 5 2 1</td>
<td>3133521</td>
</tr>
<tr>
<td>EMRSA 4</td>
<td>3 1 1 3 3 2 2</td>
<td>3113322</td>
</tr>
<tr>
<td>EMRSA 5</td>
<td>3 1 1 x 5 3 2</td>
<td>311x352</td>
</tr>
<tr>
<td>EMRSA 6</td>
<td>3 1 1 3 3 1 3</td>
<td>3133313</td>
</tr>
<tr>
<td>EMRSA 7</td>
<td>3 1 1 3 3 2 2</td>
<td>3113322</td>
</tr>
<tr>
<td>EMRSA 8</td>
<td>3 1 1 4 5 7 2</td>
<td>3114572</td>
</tr>
<tr>
<td>EMRSA 9</td>
<td>3 1 1 3 3 2 2</td>
<td>3113322</td>
</tr>
<tr>
<td>EMRSA 10</td>
<td>3 1 1 3 7 8 2</td>
<td>3113782</td>
</tr>
<tr>
<td>EMRSA 11</td>
<td>2 1 1 3 3 2 2</td>
<td>2113322</td>
</tr>
<tr>
<td>EMRSA 12</td>
<td>3 1 1 3 4 1 3</td>
<td>3133413</td>
</tr>
<tr>
<td>EMRSA 13</td>
<td>3 1 1 3 3 1 3</td>
<td>3133313</td>
</tr>
<tr>
<td>EMRSA 14</td>
<td>3 1 1 3 3 1 3</td>
<td>3133313</td>
</tr>
<tr>
<td>EMRSA 15</td>
<td>1 5 0 3 8 x 2</td>
<td>15038x2</td>
</tr>
<tr>
<td>EMRSA 16</td>
<td>1 2 2 2 5 3 2</td>
<td>1222532</td>
</tr>
</tbody>
</table>

*x, locus in which no amplification was obtained.
been described by other authors, first in eukaryotes, and subsequently in prokaryotes (Jeffreys et al., 1985; Frothingham et al., 1998). The availability of multiple sequenced \textit{S. aureus} genomes has allowed us to make a comparative study of the VNTRs, and their potential role in determining the evolution of UK hospital MRSA isolates.

Several of the eukaryotic VNTRs have been associated with important functional effects; for example, they influence the transcription of insulin, and are associated with increased risk of insulin-dependent diabetes and several cancers (Bennett et al., 1995; Krontiris et al., 1993). Within bacteria, they have been described within both coding and non-coding regions, and, in \textit{Haemophilus influenzae}, the loss or gain of repeats has been associated with the phase variation that plays an important role in its pathogenicity (Van Belkum et al., 1997). Although all of the SIRUs are in non-coding regions on the positive strand, the coding region for the protein A gene \textit{spa} is on the complementary strand of SIRU21. Protein A is a cell-wall-anchored protein, and has been shown to have a possible role in the virulence of \textit{S. aureus} infections; the effect of variable numbers of tandem repeats on the \textit{spa} gene is not known.

In this study, the repeat units of the SIRUs differed in length, varying from 48 to 159 bp. All the repeat units at the different loci contained point mutations, and small insertions and deletions; this phenomenon was first described in the minisatellite in the human insulin gene, and has since been shown to be the case in all VNTR loci described, both eukaryotic and prokaryotic (Mazars et al., 2001; Owerbach & Aagaard, 1984). Like the VNTRs described in other bacterial genomes, and unlike those in eukaryotes, the repeat units in \textit{S. aureus} are not G+C rich (Le Flèche et al., 2001).

Variations in copy number were observed in all loci, within both the sequenced \textit{S. aureus} genomes and the EMRSA strains used in the field trial. Variation in repeat number within some of the loci was greater (SIRU05, 13 and 21), and therefore demonstrated a greater discriminatory power; this is similar to the phenomenon described in \textit{M. tuberculosis} (Supply et al., 2000). At one time, the variation in copy number was thought to be due to an unequal crossing-over exchange between homologous chromosomes, but Wolff and colleagues demonstrated that there was no exchange of flanking DNA when a repeat unit was lost (Wolff et al., 1988). Other theories have involved misalignment mutational mechanisms, or changes in allele number arising from mutation initiated by double-strand breaks (Lovett, 2004; Paques et al., 1998). Also, as described in \textit{M. tuberculosis}, within some loci, certain allele copy numbers are more frequent, but the presence of all the loci adds to the discriminatory power of the technique (Mazars et al., 2001). The factors responsible for the difference in mutation rates are not clear, but mutation rates may be promoted by factors such as short repeat units, high copy number arrays and repeat sequence homogeneity (Rocha, 2003). Interestingly the loci with the longest repeat units in \textit{S. aureus} have the least variation in copy number.

Partial repeats of varying length are present at the end of all the loci; this has been described in \textit{M. tuberculosis}, where in H37Rv, for example, the ETR-B locus contains three complete copies of the 57 bp repeat, and also an additional eight bases corresponding to the beginning of another tandem repeat (Frothingham & Meeker-O’Connell, 1998). The significance of these partial repeats is not known.

The flanking regions of each individual locus had a high

Fig. 5. Dendrogram deduced from the cluster analysis of the EMRSA strains. PFGE profiles, SIRU profiles and corresponding MLST types are shown. The box highlights SIRU profiles that differ from each other by a maximum of two loci.
sequence homology with other sequenced genomes, but they did have point mutations. The availability of multiple sequenced genomes meant that our study did not suffer from the limitations of some other studies, in that we were able to design primers in a homologous region (Pourcel et al., 2001).

The epidemiology and evolution of MRSA have been studied using several different typing techniques. The most widely used epidemiological tool is PFGE, which has a high discriminatory power. However, in some instances, the genetic markers that PFGE examines undergo rapid evolutionary change, and strains that are clonally related may produce diverse banding patterns (Blanc et al., 2002). MLST examines the more slowly evolving genomic core, and is useful in the determination of evolutionary events. It has been applied to defining the evolutionary relationships of MRSA on a worldwide basis, but the applicability to local epidemiological settings has not been determined. The EMRSA isolates examined in our study were clustered into the same groups using PFGE, MLST and SIRU. PFGE showed the greatest discrimination, but SIRU typing had greater discrimination than MLST, being able to subdivide the isolates with the same MLST profiles. With its ability to cluster clonally related isolates together, but provide a greater degree of discrimination than MLST, SIRU indicates that the speed of change of VNTRs may be a more appropriate evolutionary clock for studying transmission events, as demonstrated in M. tuberculosis (Hawkey et al., 2003). Recently 15 VNTR sequences were used to examine the phylogeny of a diverse collection of geographically and temporally separated strains of M. tuberculosis (Sola et al., 2003). When compared with other molecular methods, variation in VNTR loci not only supported the accepted phylogenetic structure, but also corroborated a hypothetical evolutionary scenario for the early divergence of M. tuberculosis from other closely related species. Sabat and colleagues reported the presence of several VNTR loci in S. aureus (not MRSA), described as multiple-locus VNTR analysis (Sabat et al., 2003). The numbers of copies present were not determined, and the primer pairs were multiplexed to give a fingerprint pattern. The gene spa, encoding Protein A, contains a VNTR sequence within it, which we have identified as SIRU21. Variation in copy number and DNA sequence within the repeat units of spa has been used as a genetic marker in several studies, but it restricts the evolutionary studies to examining only one region of the genome (Koreen et al., 2004). With our description of SIRUs, we expect that use can be made of the SIRU loci for enabling phylogenetic reconstruction of S. aureus, similar to the case with VNTR sequences and M. tuberculosis.

ACKNOWLEDGEMENTS

The authors wish to thank Wyeth for the support of K. J. Hardy, the Danish Center for Scientific Computing, whose grant supported part of this work, and BBSRC grant no. 6/13209 for sequencing.

REFERENCES


Le Fleche, P., Hauck, Y., Onteniente, L. & 7 other authors (2001). A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol 1, 2.


