Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray

Julia M.-L. Sung,1,† David H. Lloyd2 and Jodi A. Lindsay1

Correspondence
Jodi A. Lindsay
jlindsay@sgul.ac.uk

1Centre for Infection, Department of Cellular & Molecular Medicine, St George’s, University of London, Cranmer Terrace, London SW17 ORE, UK
2Department of Veterinary Clinical Sciences, Royal Veterinary College, University of London, Hawkshead Campus, North Mymms AL9 7TA, UK

Received 6 December 2007
Revised 4 April 2008
Accepted 9 April 2008

Staphylococcus aureus is a commensal and pathogen of several mammalian species, particularly humans and cattle. We aimed to (i) identify S. aureus genes associated with host specificity, (ii) determine the relatedness of human and animal isolates, and (iii) identify whether human and animal isolates typically exchanged mobile genetic elements encoding virulence and resistance genes. Using a well-validated seven-strain S. aureus microarray, we compared 56 UK S. aureus isolates that caused infection in cows, horses, goats, sheep and a camel with 161 human S. aureus isolates from healthy carriers and community acquired infections in the UK. We had previously shown that human isolates are clustered into ten dominant and a few minor lineages, each with unique combinations of surface proteins predicted to bind to human proteins. We found that the animal-associated S. aureus clustered into ten lineages, with 61% assigned to four lineages, ST151, ST771, ST130 and ST873, that were unique to animals. The majority of bovine mastitis was caused by isolates of lineage ST151, ST771 and ST97, but a few human lineages also caused mastitis. S. aureus isolated from horses were more likely to cluster into human-associated lineages, with 54% of horse-associated S. aureus assigned to the human clusters CC1, CC8 and CC22; along with the presence of some multi-drug resistant strains, this suggests a human origin. This is the most comprehensive genetic comparison of human versus animal S. aureus isolates conducted, and because we used a whole-genome approach we could estimate the key genes with the greatest variability that are associated with host specificity. Several genes conserved in all human isolates were variable or missing in one or more animal lineages, including the well-characterized lineage specific genes fnbA, fnbB and coa. Interestingly, genes carried on mobile genetic elements (MGEs) such as chp, scn and sak were less common in animal S. aureus isolates, and bap was not found. There was a lot of MGE variation within lineages, and some evidence that exchange of MGEs such as bacteriophage and pathogenicity islands between animal and human lineages is feasible, but there was less evidence of antibiotic resistance gene transfer on the staphylococcal cassette chromosomes (SCC) or plasmids. Surprisingly, animal lineages are closely related to human lineages and only a handful of genes or gene combinations may be responsible for host specificity.

Abbreviations: CA-MRSA, community-acquired MRSA; CC, clonal complex; CV, core variable; MLST, multi-locus sequence typing; MRSA, meticillin-resistant Staphylococcus aureus; RVC, Royal Veterinary College; SaPI, S. aureus pathogenicity islands; SCC, staphylococcal cassette chromosome; ST, sequence type; VLA, Veterinary Laboratories Agency; VRE, vancomycin-resistant enterococci; VRSA, vancomycin-resistant S. aureus.

Microarray data from this study can be found in B@G@Sbase and ArrayExpress, with accession numbers A-BUGS-17 and E-BUGS-62. A supplementary table with details of the Staphylococcus aureus isolates studied is available with the online version of this paper.

1Present address: Division of Cell Biology and Imaging, National Institute for Biological Standards and Control, South Mimms EN6 3QG, UK.

DOI 10.1099/mic.0.2007/015289-0

Microbiology (2008), 154, 1949–1959

Printed in Great Britain
INTRODUCTION

In human and veterinary medicine, infections due to staphylococci are of major importance. In humans, Staphylococcus aureus is a persistent resident of the human nose in 20% of the healthy population, and intermittently carried by another 60% (Kluytmans et al., 1997). It is a common cause of community-acquired skin infections, and a major cause of hospital-acquired infections such as surgical and catheter-site infections, bacteremia and pneumonia. S. aureus can cause a range of infections in livestock and is the most common aetiological agent of contagious bovine mastitis, causing significant losses in the dairy industry (Sears & McCarthy, 2003).

A variety of methods, such as multilocus enzyme electrophoresis (MLEE), random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR), restriction fragment length polymorphism (RFLP), pulsed-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST) have been used in epidemiological investigations of human and farm animal staphylococci (Kapur et al., 1995; Fitzgerald et al., 1997; Zadoks et al., 2000, 2002; Smith et al., 2005; Jørgensen et al., 2005). Many of the studies are in agreement with the general concept of host specialization among S. aureus isolates. Although these techniques are useful for classifying isolates into convenient intraspecies subsets, most provide little information about estimates of overall genetic relationships among different S. aureus strains, nor do they identify exactly which genes are associated with specific host–pathogen interactions. Gene products associated with host specificity would be ideal targets for therapeutic agents to prevent carriage and infection in humans or animals, reducing morbidity, mortality and agricultural losses.

We have recently developed and validated a seven-strain S. aureus whole-genome microarray (Witney et al., 2005) and used it to investigate the population structure of human S. aureus (Lindsay et al., 2006). Ten human lineages of S. aureus dominate, with some minor lineages. Each lineage has a unique combination of surface proteins and regulators [called core variable (CV) genes], and corresponds to the clonal complexes (CCs) identified by MLST typing. Each lineage has not evolved independently, but is the result of multiple recombinations of CV genes between lineages. Furthermore, all strains carry a variety of mobile genetic elements (MGE) encoding resistance and putative virulence genes, and there is much variation within lineages indicating frequent horizontal transfer.

The emergence and dissemination of antimicrobial resistance amongst staphylococci is an important problem in clinical settings, and in particular, the increasing prevalence of meticillin-resistant S. aureus (MRSA) is of concern to hospitals worldwide (Ayliffe 1997; Gould 2005). MRSA is also of concern in animals such as dogs, horses and pigs (Loeffler et al., 2005; Weese et al., 2005; Huijsdens et al., 2006). These animals could act as a reservoir of human isolates (this is likely for dogs), or could represent an opportunity for the evolution of increasingly pathogenic and resistant strains. Resistance genes are typically encoded on MGEs such as staphylococcal cassette chromosome (SCC) mec, plasmids and transposons (Lindsay & Holden, 2006). The emergence of increasingly virulent strains of S. aureus in humans is also being reported, such as the community-acquired (CA-)MRSA that produces Panton–Valentine leukocidin (PVL) toxin. PVL (encoded by PVL-luk) and other known toxins such as toxic shock syndrome toxin (tst) and enterotoxin A (eta) are encoded on MGEs such as bacteriophage and S. aureus pathogenicity islands (SaPI), and also transfer horizontally between S. aureus isolates, leading to the emergence of new clones with novel clinical implications (Lindsay & Holden, 2006; Waldron & Lindsay, 2006). The biggest concern is the potential transfer of vancomycin resistance (vanA) from vancomycin-resistant enterococci (VRE) to human S. aureus; since VRE are found in the agricultural setting (Witte, 2000), this is a potential locale for the emergence of fully vancomycin-resistant S. aureus (VRSA), which would be a major catastrophe for human healthcare.

In this study, we carried out comparative genomics using a well-validated, seven-strain S. aureus microarray in order to comprehensively screen for genes associated with human host specificity, gain a better understanding of the genetic relatedness within natural populations of animal and human S. aureus, and determine how frequently they might exchange DNA.

METHODS

Bacterial isolates and species identification. A total of 55 farm-animal-associated S. aureus were analysed; these are detailed in Supplementary Table S1, available with the online version of this paper. Animal isolates were from cows (n=37, predominantly mastitis), horses (n=13), sheep (n=2), goats (n=2) and a camel (n=1). The isolates from animals were from different regions of the UK and submitted by veterinary practitioners to the microbiology diagnostic laboratory at the Royal Veterinary College (RVC) or the Veterinary Laboratory Agency (VLA; kindly donated by Chris Teale). In addition, a bovine mastitis isolate from Ireland (RF122, also known as ET3-1) was recently sequenced by Herron-Olsen and colleagues (Herron-Olsen et al., 2007) and kindly sent to us by Ross Fitzgerald (GenBank AJ938182). Isolates were identified as S. aureus by growth on mannitol salt agar with yellow pigmentation, and catalase-positive and staphyloase-positive reactions (erythrocyte agglutination for the detection of clumping factor). To distinguish S. aureus from the typical animal pathogen Staphylococcus intermedius, they were also tested for the presence of nuc and fem genes by PCR, according to the method described by Becker et al. (2005). Antimicrobial resistance was determined using standard disk diffusion methods. The previously described 161 human S. aureus isolates were from a wide range of sources, representing the major dominant lineages of S. aureus from hospitals and the community, including hospital MRSA (Feil et al., 2003; Lindsay et al., 2006).

Microarrays. The S. aureus microarray employed in this study has been described and comprehensively validated previously (Witney et al., 2005; Lindsay et al., 2006). The array design is available in BugSetSbase (accession number: A-BUGS-17; http://bugs.sgul.ac.uk/A-BUGS-17) and also ArrayExpress (accession number: A-BUGS-17).
Briefly, it contains 3,623 PCR products representing every predicted ORF in the first seven complete *S. aureus* genome sequencing projects. These sequenced strains are all from humans and correspond to lineages CC1, CC5, CC8 and CC30. The microarray was also designed to include spots corresponding to putative target recognition domain (TRD) regions of *sau1hsd* that show major sequence variation, as described by Waldron & Lindsay (2006). PCR products were printed in duplicate on GAP5 slides (Corning). DNA was fixed using UV light and blocked with BSA prior to hybridization. All human strains had previously been hybridized to the microarray and reported by Lindsay et al. (2006).

**DNA isolation, labelling and hybridization.** DNA was extracted using genomic-tip 100/G columns (Qiagen) or Bacterial Genomic Prep (Edge Biosystems), and concentration was measured as A260 using the Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies). Bulk reference DNA was prepared from MRS252 using either the caesium chloride method (Lindsay et al., 1998) or with Bacterial Genomic Prep. Four micrograms of test DNA was labelled using Cy3 dye and DNA polymerase I large fragment (Klenow; Invitrogen), and 4 μg of reference DNA was labelled using Cy5 dye. The two samples were pooled and hybridized to a *S. aureus* microarray overnight, prior to washing and scanning (Lindsay et al., 2006).

**Data generation and analysis.** Microarrays were scanned using an Affymetrix 428 scanner and images were converted to raw data by BlueFuse for Microarrays 2.0 (BlueGnome, Cambridge, UK). Data analysis was performed in GeneSpring 7.0 (Agilent). The presence or absence of a gene or gene variant region in a test isolate was determined by its hybridization to the relevant spot (detected by Cy3 fluorescence) in comparison to hybridization of the reference strain MRS252 (detected by Cy5 fluorescence), and was expressed as a ratio. This methodology was fully validated using the seven sequenced strains, which generated microarray data that correlated well with expected gene presence or absence, even for those genes not found in the reference strain MRS252 (Witney et al., 2005). Please refer to Whitney et al. (2005) for a fuller explanation of the limitations of using microarrays to define a gene as present or absent, particularly in cases where there is gene variation. Fully annotated microarray data are deposited in the public databases BGC@SBase (accession number: E-BUGS-62; http://bugs.sgl.ac.uk/E-BUGS-62) and also ArrayExpress (accession number: E-BUGS-62).

Raw data from the previously reported human *S. aureus* isolate study (Lindsay et al., 2006) and from the new *S. aureus* animal isolates were analyzed in GeneSpring using locally weighted scatterplot smoothing (LOWESS), with 50% of the data used for smoothing and a control channel cutoff of 0.01. Condition tree clustering using the Spearman correlation was performed as a function in GeneSpring, according to Lindsay et al. (2006). Briefly, this involved using the set of 728 CV genes that vary according to lineage, and clustering these isolates into lineages using the Spearman correlation. Secondly, strains were inspected for the presence of MGEs using virtual genomes artificially constructed in GeneSpring from known MGE genes in order from the seven sequenced strains (Witney et al., 2005). MGEs (including SCCmec, bacteriophage, SaPI, plasmids and transposons) are present when all or most genes in an MGE are present (by ratio) in a given strain, although it is generally more common to find several neighbouring genes present that represent a mosaic fragment of an MGE that has probably recombined into a new MGE element (Lindsay et al., 2006; Lindsay & Holden, 2006).

**Epidemiological typing.** The lineages of the animal *S. aureus* were identified by clustering the isolates using 728 CV genes (non-core, non-MGE), according to the method of Lindsay et al. (2006). That study showed that clusters generated by microarray analysis agreed with the division of isolates into MLST CCs, confirming the usefulness of both methods. Representative isolates of each cluster or lineage identified by microarray were thus subjected to MLST typing in order to assign a sequence type (ST) and CC number to each group (Enright et al., 2000). In brief, seven housekeeping genes were sequenced and each allele assigned a profile, with the combined profiles generating a ST assigned by the MLST database (http://www.mlst.net). STs were clustered into CCs using eBURST, employing the stringent (conservative) group definition in which all members assigned to the same group share at least six of the seven loci with at least one other member of the group (Feil et al., 2004).

**Sequencing of sau1hsds.** *sau1hsd* genes sequences from isolates representing lineages ST151 (CC151), ST771 (CC133), ST130 (CC130), ST188 (CC188), ST1 (CC1), ST73 (CC873) and ST97 (CC97) were sequenced. PCR products were amplified using Platinum Pfx DNA polymerase (Invitrogen) with conditions of 94°C for 5 min followed by 25 cycles of 94°C for 30 s, 52°C for 30 s, 68°C for 140 s. Products were cleaned on QiaQuick spin columns (Qiagen), and both strands were sequenced by Lark Technologies. The primers used to amplify and sequence the *hsd* genes were 5'-AATGTACATCCATGAAACATTGG-3' and 5'-GAC-ACTGCGCTTCACTGCCC-3' for *sau1hsdS*, and 5'-ATGCATA-CCTGGAAGACCTGG-3' and 5'-CAATTAATAGGTGTATCA-GG-3' for *sau1hsdS2*.

**Comparison of microarray data and PCR for detection of host-specific genes.** To support our microarray evidence, we used PCR to show whether the presence or absence of *sca*, *chip*, *sak* or *sea* was associated with host specificity, and whether the 1,500 bp insert was missing from the 3' end of *fmtB* in lineage ST151. We used the primers in Table 2, which are the same ones used to generate the PCR products spotted on our microarray. All 56 animal isolates (including the sequenced RF122 used as a control) and 35 representative human invasive disease isolates were tested. PCR products were amplified using HotStar Taq polymerase (Qiagen), with conditions of 94°C for 30 s, 55°C for 140 s, followed by 25 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 120 s. A 5 μl sample of each reaction was run on a 1% agarose gel, and a reaction was recorded as positive if a single clear band of the correct size was present.

**RESULTS**

**Distribution of animal lineages**

Using the seven-strain *S. aureus* microarray, the 56 animal isolates were clustered into ten different lineages (Fig. 1). MLST typing and eBURST analysis of representatives of each lineage assigned them to the STs shown in Table 1. Most animal isolates belonged to animal lineages ST151, ST771, ST130 and ST873. Some animal isolates belonged to lineages ST97 and ST188, which have previously been described in humans at low frequency. A few animal isolates belonged to the dominant human clusters CC1, CC8 and CC22.

Of the 37 bovine isolates, 81% were from three dominant lineages, ST151 (*n* = 13), ST771 (*n* = 9) and ST97 (*n* = 8). ST151 has previously been described as the major lineage of *S. aureus* associated with bovine mastitis in the UK, but not in the USA or Chile (Smith et al., 2005). The recently sequenced Irish isolate, RF122, belongs to ST151. As of November 2007, the MLST database of approximately...
1500 isolates did not describe the novel ST, ST771 (http://www.mlst.net). ST97 was described as a common bovine mastitis lineage in the US and the most common in Chile but was not previously described in the UK (Smith et al., 2005). In our collection, the ST97 isolates were more drug resistant than other animal isolates (Table S1). Interestingly, all of the ST97 isolates in our study came from the VLA, and were from outside of the RVC catchment area around Essex, UK. However isolates of lineages isolated in the Essex region (RVC collection) were also found throughout the UK. The remaining bovine isolates were assigned to ST130 (n=2), reported previously in cows and a goat from Norway (Jørgensen et al., 2005); ST188 (n=3), found in humans at low frequency (Feil et al., 2003); ST1 (n=1) a dominant human lineage (Feil et al., 2003); and one orphan.

Of the 13 isolates from horses, more than half (n=7, 54%) were assigned to the dominant human lineages ST1, ST22 and ST8, and several were multi-drug resistant. The ST22 lineage includes EMRSA-15, one of the two major clones of UK epidemic hospital-associated MRSA. However, the equine ST22 in this study is phenotypically meticillin-sensitive. ST8 is a common human meticillin-sensitive S. aureus (http://www.mlst.net).

**Fig. 1.** Microarray data from 217 S. aureus (161 human carriage and invasive isolates and 56 animal isolates), clustered by Spearman correlation using 728 CV genes. Vertical lines represent each isolate and are coloured in the tree and upper row of artificially coloured blocks below the microarray by lineage. Lineage numbers are printed underneath. The lower row of artificially coloured blocks are coloured pale green for cattle isolates and pale blue for horse isolates; human isolates are coloured dark blue (invasive) or purple (carriage). Within the main picture, the horizontal lines represent individual genes that are listed in order in Fig. 2. A yellow or red signal indicates a gene is present; a blue, white or grey signal indicates it is absent. Note that the animal lineages do not cluster separately from the human lineages.
**Table 1. Distribution of each lineage in 56 animal and 161 human isolates**

Numbers in parentheses show the percentage of the total number of isolates. Human isolate data are from the study by Lindsay et al. (2006).

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Animal</th>
<th>Human*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>CC</td>
<td>Bovine</td>
</tr>
<tr>
<td>151</td>
<td>151</td>
<td>13 (35)</td>
</tr>
<tr>
<td>771</td>
<td>133</td>
<td>9 (24)</td>
</tr>
<tr>
<td>97</td>
<td>97</td>
<td>8 (22)</td>
</tr>
<tr>
<td>130</td>
<td>130</td>
<td>2 (5)</td>
</tr>
<tr>
<td>188†</td>
<td>188</td>
<td>3 (8)</td>
</tr>
<tr>
<td>1‡</td>
<td>1</td>
<td>1 (3)</td>
</tr>
<tr>
<td>873</td>
<td>873</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Orphan</td>
<td>1</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>13</td>
</tr>
</tbody>
</table>

*Human isolates grouped into major lineages (CC) only.
†ST188 and ST1 are clustered into the same lineage by MLST using seven housekeeping genes. By microarray analysis of 728 genes, these two types belong to separate lineages.

**Table 2. Distribution of scn, chp, sak and sea genes in human versus animal isolates by microarray**

The animal group (n=56) includes the 14 ST151 isolates. Genes were regarded as present if the normalized ratio of test strain signal divided by reference strain signal was greater than 0.5, except for sea, where the ratio used was 0.75 (the percentage of isolates which tested positive to each gene is given in parentheses).

<table>
<thead>
<tr>
<th>Primers (5’–3’)</th>
<th>Probe size (nt)</th>
<th>Human (n=161)</th>
<th>Animal (n=56)</th>
<th>ST151 (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR2035 (scn)</td>
<td>ATACTTGCGGGAACCTTTAGCAA</td>
<td>320</td>
<td>154 (96 %)</td>
<td>12 (21 %)</td>
</tr>
<tr>
<td></td>
<td>TTTTATGCTTCTGTAATTTCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR2036 (chp)</td>
<td>TTTTTAAACGCAGGGAGATCATGTA</td>
<td>404</td>
<td>134 (83 %)</td>
<td>8 (14 %)</td>
</tr>
<tr>
<td></td>
<td>TGCATATCAGATTTTTTTCCAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR2039 (sak)</td>
<td>TGAGTTAAGTGCATCAAGTTCA</td>
<td>403</td>
<td>134 (83 %)</td>
<td>11 (20 %)</td>
</tr>
<tr>
<td></td>
<td>CTTTTGTAATTAAGTTGAATCCAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR2043 (sea)</td>
<td>CTTTTATCATCATGCCCTIAACGTG</td>
<td>721</td>
<td>41 (25 %)</td>
<td>8 (14 %)</td>
</tr>
<tr>
<td></td>
<td>TGCACTTTTCTCAGATTAAATCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR2248 (fmtB)</td>
<td>AATGGAGATCCGGCAATCATGTTG</td>
<td>725</td>
<td>161 (100 %)</td>
<td>43 (77 %)</td>
</tr>
<tr>
<td></td>
<td>CATCATTATTTGTTGGTGTA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

–, Analysis not performed.
of previously described CV (human *S. aureus*) genes, and there was no evidence that variation between each of the animal lineages was greater than variation between animal and human lineages (Fig. 1).

To further explore the association of the genes with host specificity, we looked for genes found in human isolates (core genes, found in \(>95\%\)), which were not present in animal isolates. A visual inspection of 2013 core genes found no genes in all human isolates that were uniformly absent in animal isolates. The ‘Find Significant Parameter’ functions available in GeneSpring can be adapted to identify gene differences associated with parameters such as source of isolate; this also identified no genes. Several genes that were under-represented, but not uniformly missing, in isolates of animal origin were identified, and these genes were typically associated with lineage (thus they would be newly classified as CV genes). They include SAR0179 (putative transporter similar to LmrP, missing in ST151, ST771, ST130 and some ST97), SAR0744 (putative DNA photolyase, missing in ST151 and ST771), SAR0745 and SAR0746 (putative membrane proteins missing in ST151, ST771, ST130, ST873), SAR2257 (multi-drug resistance transport, missing in ST151, ST771, ST130, ST873), SAR2518 and SAR2519 (ATP binding, missing in ST151 and some ST97), SAR2564 (membrane protein, missing in ST151, ST771, ST97 and ST188) and SAR2633 (regulator of TetR family, missing in ST151, ST771, ST130, ST873).

The 728 human CV genes associated with specific human lineages were also analysed to determine if any were relatively more or less common in animal lineages. Examples of the distribution of genes are shown in Fig. 2. There was a surprisingly high number of human CV genes found in animal lineages, consistent with the fact that

---

**Fig. 2.** CV genes studied in Fig. 1. Genes are listed in order by name and by their identifier or annotated gene number (R, MRSA252; N, N315; B, B325; M, MW2). ‘v’ or * denotes a PCR product designed to a specific variant region. A black box indicates the gene is present in that CC or ST. ‘u’ indicates variation in gene distribution for that lineage.
animal isolates did not cluster separately from human isolates using the microarray. A few genes were less common in animal isolates, including SAR2363 (membrane protein, missing in ST151, ST771, ST97 and ST188), SAR1329-32 (two-component signal transduction missing in ST151), SAR1223 (LytN, missing in ST151, ST771, ST873 and ST97) and SAR1224 (FmhC, missing in ST151, ST873 and ST97). There were differences in several well-characterized surface proteins. fnbA or fnbB were found in all the animal isolates, but the four types of variable region on the microarray (corresponding to the four sequenced human lineages) were not found in the animal isolates, suggesting these isolates have unique variable regions. This is confirmed by the RF122 sequencing project (Herron-Olsen et al., 2007), which shows a unique variable region in the fnbA and fnbB genes. Similarly, nearly all strains had a coagulase gene, but the four types of variable region of the coagulase gene on the microarray were not found in the animal isolates. Again, this was confirmed by the RF122 sequencing project, as this strain has a unique coagulase variable region. The ebb gene fragment on the microarray hybridized with isolates from only two human lineages, and no animal lineages, indeed the RF122 sequencing project suggests that this gene is broken into three smaller genes and a large region is completely missing. The vra genes, possibly associated with cell-wall synthesis and resistance to vancomycin, were found in three human genes and a large region is completely missing. The vra genes, possibly associated with cell-wall synthesis and resistance to vancomycin, were found in three human lineages and no animal lineages. As our microarray was designed using only genes from seven human S. aureus strains, we were unable to detect animal S. aureus specific genes.

Curiously, our microarray analysis demonstrated that fmtB, a gene encoding a cell-wall-associated protein, has a large fragment missing from all ST151 isolates. Tn551-mediated insertional inactivation of the fmtB gene has been shown to drastically reduce resistance of S. aureus to β-lactam antibiotics such as meticillin, cefoxitin and imipenem, although susceptibility of non-β-lactam inhibitors (e.g. fosfomycin, bacitracin and vancomycin) and protein synthesis inhibitors (e.g. chloramphenicol and tetracycline) were not affected (Komatsuzaawa et al., 2000). It is notable that most of the animal isolates were more sensitive to antibiotics than human isolates and surprisingly, two isolates that carried SCCmec elements and were positive for mecA were phenotypically meticillin sensitive (see Table S1).

Sau1 restriction–modification system

Using microarrays, the sau1hsdS1 gene variant regions of the animal isolates appeared to be different to previously described S. aureus isolates, and to be distributed according to lineage. This confirmed our earlier hypothesis that each lineage has a unique sau1hsdS gene that encodes Sau1 restriction–modification specificity, which then contributes to the relatively independent evolution of each lineage (see Discussion) (Waldron & Lindsay, 2006). Sequencing revealed that sau1hsdS1 genes of ST151 (CC151), ST771 (CC133) and ST130 (CC130) have similar conserved regions to the human isolates, mainly at the short 5′ end known otherwise as the proximal conserved region, but had novel variable regions predicted to code the binding region to specific DNA sequences (data not shown). This is exactly as predicted and supports the hypothesis that independent lineages have unique sau1hsdS genes. The sau1hsdS1 genes of ST188 and ST1 from animals, both of which are clustered into the same lineage by MLST but not microarray, are homologous to the sau1hsdS1 genes of human CC1. At this stage, we have been unable to amplify a fragment of the ST188 sau1hsdS2 gene for sequencing, suggesting it is different to that found in CC1. This supports our previous hypothesis (Lindsay et al., 2006) that ST188 isolates have relatively recently diverged from ST1 isolates and that while there are multiple differences between the two lineages, this is not yet detected in the seven housekeeping genes used for MLST analysis.

Distribution of mobile genetic elements (MGEs)

In human S. aureus, MGEs which include plasmids, bacteriophage, SaPI, transposons and SCCs (including SCCmec), vary significantly (Lindsay et al., 2006). Some stable MGE are common to many isolates or within some lineages, whilst others vary substantially, even between isolates of the same lineage, indicating frequent transfer and instability. Each MGE has a mosaic structure, and is made up of a combination of fragments that may be found in other MGEs, and recombination between MGEs is high. Since microarrays do not provide information about the position of genes in the chromosome, we used ‘composite genomes’ – artificial groupings of microarray data in the gene order found in known MGE – to estimate their location based on the sequenced MGE. These are described in Witney et al. (2005) and all the composite genomes are available in the deposited databases.

The data for the animal-associated S. aureus suggest that animal lineages have SaPI and bacteriophage genes that are similar to those found in human isolates. That is, genes homologous enough to generate a signal on the human S. aureus microarray were identified, and on many occasions, multiple genes known to be next to each other in human S. aureus MGE were also found in animal S. aureus. Distribution of genes shows variation between all isolates, including within lineages, but there is still a trend towards increased conservation within lineages. Notably, animal isolates were significantly less likely to carry some of the ‘immune evasion cluster’ genes encoded on β-haemolysin converting bacteriophage of phage group 3 (van Wamel et al. 2006), including chp, scin and sak (Tables 2 and 3). The presence or absence of these genes was not lineage dependent. Monecke et al. (2007) and Kumagai et al. (2007) have already reported a low incidence of these genes in bovine isolates of S. aureus.
SCC elements were only found in the equine CC22 isolate and one equine CC8 isolate, suggesting no transfer of these elements to animal lineages. Human plasmids were rare in animal isolates, although the transposon encoding β-lactamase (commonly found on plasmids and very common in human isolates) was found only in some ST97 isolates, as were cadDX genes. No other transposons were found. These data suggest that bacteriophage and SaPI can transfer between animal and human isolates, but relatively infrequently, while plasmids, SCC and transposons transfer very infrequently.

It has been suggested that animal isolates preferentially carry a gene called bap, associated with biofilm production, on a SaPI (Tormo et al., 2005). Using the PCR assay described by Tormo et al. (2005) and primers sasp-6m and sasp-7c, we were unable to identify this gene in any of our isolates (data not shown). Our positive control was staphylococcal strain 219R, isolated from a cow, which generated a band of the correct size with these primers, and when sequenced showed 98% homology to the Staphylococcus xylosus bap gene. We note the low incidence of the bap gene in another animal strain collection (Vautor et al., 2008).

### Comparison of microarray data and PCR

The PCR data (Table 3) confirm the microarray data, showing an association of scn, chp and sak with human versus animal isolates, and confirm that the deletion in the fmtB gene is associated with ST151 isolates. There are some discrepancies between the microarray determination of present or absent based on a cut-off ratio, compared to the PCR method. This could be due to the fact that microarrays detect homology over the whole length of the probe, which may be hundreds of base pairs long, whereas PCR detects homology to two short sequences that are spaced at a conserved distance from each other. Microarrays can be influenced by the presence of homologous genes, which may have contributed to the difficulty in detecting sea which belongs to a family of homologous superantigen genes that vary considerably between strains. However, in our previous studies, we showed that microarray data compared well to whole genome sequence data (Witney et al., 2005). The advantage to microarray data is that internal controls for most genes are present and all gene spots are related to all other gene spots for the strain being tested, which effectively act as further controls to identify poor-quality data. We have also previously shown that microarray data often compare well to previously generated PCR data (Peacock et al., 2002; Lindsay et al., 2006), but that some individual PCR primer pairs perform poorly in PCR reactions. Validation of individual PCR primer pairs is not common, although increasing numbers of strains are being sequenced, which will make it possible in the future.

### DISCUSSION

The data obtained from this study showed that the majority of the animal-associated S. aureus clustered into animal-specific lineages or in predominantly animal lineages. This was particularly the case for bovine mastitis isolates and those associated with goats and sheep (88%). However, some of the lineages had not been described in previous MLST studies of bovine S. aureus in the UK, USA, Chile and Norway (Smith et al., 2005; Jorgensen et al., 2005). We also noted substantial geographical variation in the distribution of isolates both within the UK (Table S1), and between reported lineages in other countries. Further work will be necessary to determine if this represents geographical variation, or associations with particular types of cattle or practice.

Interestingly, 12 bovine mastitis cases were caused by isolates from lineages previously described in humans. One case involved the major human lineage CC1. Recently, transmission of MRSA CC1 lineage isolates between cows and humans has been reported (Juha´sz-Kaszanyitzky et al., 2007). Three cases involved ST188, which is relatively uncommon in humans but has not been reported in animals before. Eight cases involved CC97, also relatively

| Table 3. Comparison of microarray and PCR detection of scn, chp, sak and sea genes and the fmtB insert region |
|-----------------|----------------|----------------|
|                 | Human (n=35)   | Animal (n=56)  | ST151 (n=14) |
| **Array**       | **PCR**        | **Discrepant** | **Array**    | **PCR**        | **Discrepant** | **Array**    | **PCR**        | **Discrepant** |
| scn             | 35 (100)       | 34 (97)        | 1 (3)        | 12 (21)       | 6 (11)         | 10 (18)       | –            | –            | –            |
| chp             | 30 (86)        | 27 (77)        | 3 (9)        | 8 (14)        | 4 (7)          | 11 (20)       | –            | –            | –            |
| sak             | 32 (91)        | 30 (86)        | 2 (6)        | 11 (20)       | 6 (11)         | 9 (16)        | –            | –            | –            |
| sea             | 10 (29)        | 11 (31)        | 9 (26)       | 8 (14)        | 2 (4)          | 8 (14)        | –            | –            | –            |
| fmtB            | 35 (100)       | 33 (94)        | 2 (6)        | 43 (77)       | 41 (73)        | 4 (7)         | 1 (7)        | 2 (14)       | 1 (7)        |

–, Analysis not performed.
rare in humans. This argues against the idea that only animal isolates cause animal disease and only human isolates are carried by humans or cause human disease. It is not known if the humans previously identified as carrying ST97 had contact with farming or agriculture. It is tempting to speculate that disease can be caused by any

S. aureus, and that opportunity to cause disease is the key. Thus, the animal isolates may be more prevalent on animals in the dairy environment or on those which carry S. aureus on the bovine teat, skin and nares (Sears & McCarthy, 2003).

Horses were much more likely to be infected with isolates from human lineages, which may also reflect opportunity. There is limited evidence that horses carry S. aureus in their nares for long periods, and perhaps they are only transient carriers, as is likely for pets such as dogs. Since horses that receive veterinary care for S. aureus infection are likely to be in close contact with humans, such as during grooming and riding, this could be a source of the infecting strain. It is interesting that horses are also infected with isolates of S. aureus found in cows and other ruminants, and again, this may reflect opportunity if horses are kept on farms. It would be interesting to determine which lineages are carried by humans in close contact with horses (and/or cows) and if they persist or are transitory. The implications of humans passing S. aureus to horses and other pets are important because carriage S. aureus in healthy people are becoming increasingly virulent and resistant (Dufour et al., 2002; Gopal Rao et al., 2007). This leads to animal infections that require more complex antibiotic regimes (Loeffler et al., 2005, Weese & McCarthy, 2005), and in return, animal contact could become a potential risk to immuno-compromised humans.

Are human and animal S. aureus different to each other? Our data suggest that S. aureus from humans belong to multiple lineages, while animal S. aureus belong to a range of different lineages. Each lineage carries a unique combination of CV genes, predominantly encoding surface proteins or structures and their regulators. Surprisingly, the animal S. aureus lineages carried a wide range of CV genes previously identified in human S. aureus. The phylogenetic structure of whole S. aureus populations is often represented as a tree (including by the clustering method of GeneSpring used here, which artificially draws a tree structure) but the relationships between lineages would be more truly represented by a complex network of widespread recombination. This is clearly seen by looking at the data in Figs 1 and 2. For example, if all the lineages were classified by trees, firstly into capsule types 5 and 8 and subsequently into their agr types, the tree generated would be very different to that if strains were classified by agr type and then capsule type. So the lineages have not evolved in a way that is modelled by a tree suggesting single or a few variations leading to new branches, followed by divergent evolution. Instead, the S. aureus lineages have evolved by horizontal recombination across all lineages, and our data show that this includes the animal lineages. Animal lineages carry many CV markers found in sequenced human lineages, including all of the human capsule and agr types. This suggests that they have evolved alongside the human lineages, and they are not on a separate ‘phylogenetic tree branch’. In other words, all the lineages varied greatly, but different animal lineages were no more similar to each other than to any human lineages.

Are there human and animal specific genes accounting for host specificity? This has turned out to be a complex question, as most variation was found to be lineage specific rather than host specific. Nevertheless, a number of genes were identified that showed a trend for being more common in human isolates than animal isolates; it is likely that the most significant of these are the known host-binding protein variants fibA, coagulase and efb. However, it also seems likely that minor variations in other surface proteins are important, but they are too small to be detected by microarray. The low incidence of scn, chp and sak genes in animal isolates, typically found on a common MGE bacteriophage and implicated in human host immune evasion, may also be very important. scn (staphylococcal complement inhibitor) encodes an inhibitor of complement formation that reduces phagocytosis by neutrophils. Rooijackers et al. (2005) showed that scn is human specific and does not affect cow, goat or sheep complement. chp (chemotaxis inhibitory protein) encodes a modulator of chemokine responses that prevents neutrophil chemotaxis and activation. sak (staphylokinase) encodes an anti-opsonin and inhibitor of defensins. It is possible that all of these proteins target human immune responses specifically and not animal responses (van Wamel et al., 2006). If so, it is also possible that genes targeting animal immune systems specifically are found in animal strains, but we do not have probes for them on the microarray. Our results suggest that there may be relatively few conserved differences between S. aureus from humans and animals, and fewer than had been previously suggested. This is despite the fact that most bovine and human S. aureus appear relatively host specific for colonization and pathogenicity (Kapur et al., 1995; Zadoks et al., 2002, Jorgensen et al., 2005).

The microarray allowed us the opportunity to investigate MGE variation and exchange amongst S. aureus, which is likely to be the mechanism for the emergence of increasingly virulent and resistant strains of S. aureus, such as MRSA, CA-MRSA and VRSA (Lindsay & Holden, 2006). In the animal isolates, there were few cases of SCC, plasmids or transposons typical of human isolates. However, there was evidence of bacteriophage and SaPI carriage in animal isolates, and substantial variation in their carriage between isolates, including within lineages. SCC, plasmids and transposons usually encode resistance genes, while bacteriophage and SaPI usually encode virulence genes. This suggests MGE transfer between human and animal isolates is feasible, but probably does not occur frequently. In support of this, antibiotic resistance profiles of the animal isolates showed little
evidence of widespread antibiotic resistance in the animal lineages, although this was seen in many human lineages associated with animals (Table S1). However, it is possible that our microarray study has underestimated gene exchange, as it only contains a selection of human \textit{S. aureus} MGE genes, and not animal \textit{S. aureus} MGE genes. It could be that the major variation between animal and human \textit{S. aureus} will be in their MGE carriage.

We have recently reported a restriction–modification system in \textit{S. aureus}, Sau1 (or SauI1) that controls MGE transfer between isolates of different lineages (Waldron & Lindsay, 2006). Sau1 encodes a restriction enzyme that recognizes un-modified foreign DNA and digests it, reducing the frequency of horizontal transfer of MGE, which is likely to contribute to the independent evolution of lineages. Sau1 also encodes a modification enzyme that modifies self DNA, thus protecting it. Specificity of the Sau1 system is determined by sau1hsdS1 and sau1hsdS2 genes, which vary according to lineage. Each animal lineage tested in this study had unique specificity genes, and so each animal lineage is predicted to exchange DNA with other lineages at low frequency. This system is likely to be preventing the transfer of resistance genes to animal isolates in hospitals would be a financial and humanitarian disaster for our healthcare systems.

**ACKNOWLEDGEMENTS**

We are grateful to Jason Hind, Kate Gould, Lucy Brooks, Adam Witney and Phil Butcher from the Bacterial Microarray Group at St George’s (BuGe®; www.bugs.sgu.ac.uk), funded by The Wellcome Trust, for assistance with all microarray studies. We thank all those at the RVC and VLA involved in collecting and archiving strains. This study was supported by a grant to J.A.L. from the Department of Environment, Food and Rural Affairs, UK.

**REFERENCES**


Edited by: T. Msadek

http://mic.sgmjournals.org