The Distribution of Mobile Genetic Elements (MGEs) in MRSA CC398 Is Associated with Both Host and Country

Alex J. McCarthy1,*, Adam A. Witney2, Katherine A. Gould2, Arshnee Moodley3, Luca Guardabassi3, Andreas Voss4, Olivier Denis5, Els M. Broens6,7, Jason Hinds2, and Jodi A. Lindsay1

1Centre for Infection, Department of Clinical Sciences, St George’s University of London, United Kingdom
2Bacterial Microarray Group, Department of Clinical Sciences, St George’s University of London, United Kingdom
3Department of Disease Biology, Faculty of Life Sciences, University of Copenhagen, Stigbøjlen, Frederiksberg C, Denmark
4Department of Medical Microbiology and Infection Control, Radboud University Nijmegen Medical Centre and Canisius–Wilhelmina Hospital, Nijmegen, The Netherlands
5Laboratoire de Référence MRSA-Staphylocoques, Service de Microbiologie, Université Libre de Bruxelles, Hôpital Erasme, Brussels, Belgium
6Quantitative Veterinary Epidemiology Group, Wageningen Institute of Animal Sciences, Wageningen University, The Netherlands
7Centre for Infectious Disease Control Netherlands, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

*Corresponding author: E-mail: amccarth@sgul.ac.uk.

Accepted: 7 September 2011

Data deposition: The microarray design is available in Btg@Sbase (accession no. A-BUGS-38; http://bugs.sgul.ac.uk/A-BUGS-38) and ArrayExpress (accession no. A-BUGS-38). Fully annotated microarray data have been deposited in Btg@Sbase (accession No: E-BUGS-120; http://bugs.sgul.ac.uk/E-BUGS-120) and also ArrayExpress (accession no: E-BUGS-120).

Abstract

Methicillin-resistant Staphylococcus aureus clonal complex (CC) 398 has emerged from pigs to cause human infections in Europe and North America. We used a new 62-strain S. aureus microarray (SAM-62) to compare genomes of isolates from three geographical areas (Belgium, Denmark, and Netherlands) to understand how CC398 colonizes different mammalian hosts. The core genomes of 44 pig isolates and 32 isolates from humans did not vary. However, mobile genetic element (MGE) distribution was variable including SCCmec. q3 bacteriophage and human specificity genes (chp, sak, scn) were found in invasive human but not pig isolates. SaPI5 and putative ruminant specificity gene variants (vwb and scn) were common but not pig specific. Virulence and resistance gene carriage was host associated but country specific. We conclude MGE exchange is frequent in CC398 and greatest among populations in close contact. This feature may help determine epidemiological associations among isolates of the same lineage.

Key words: methicillin-resistant Staphylococcus aureus, host specificity, mobile genetic elements, zoonoses, bacteriophages.

Introduction

Staphylococcus aureus colonizes 25% of healthy humans and is a frequent cause of skin infections as well as invasive infections in the immunocompromised. Methicillin-resistant S. aureus (MRSA) from hospital-acquired and community-acquired reservoirs are resistant to antibiotics commonly prescribed to prevent or treat infections. In animals, S. aureus is a major cause of ruminant mastitis and also causes companion animal infections (Loeffler et al. 2005; Huijsdens et al. 2006). The use of antibiotics in farming and veterinary practice selects for MRSA and other drug-resistant S. aureus that have zoonotic potential (Wulf and Voss 2008).

Of concern is the recent zoonotic emergence of livestock-associated MRSA CC398 from pigs. MRSA CC398 colonizes pigs but rarely causes infection. Humans in contact with pigs are highly likely to be colonized (Lewis et al. 2008; Denis et al. 2009). Invasive human infections caused by MRSA CC398 were first detected in pig farmers in Europe in 2004 (Huijsdens et al. 2006) and have since spread across Europe with the highest numbers of cases in Austria, Belgium, Denmark, and The Netherlands (Huijsdens et al. 2006; Witte et al. 2007; van Cleef...
et al. 2011). MRSA CC398 has also emerged rapidly to colonize or cause human infections in North America and Asia (Khanna et al. 2008; Bhat et al. 2009; Smith et al. 2009). Worryingly, virulence genes such as Panton–Valentine leukocidin (PVL) encoded on a bacteriophage have been identified in some CC398 MRSA, and colonization and infections in humans with no known contact with animals have been reported in the United States and Europe (van Belkum et al. 2008; Welinder-Olsson et al. 2008; Wulf et al. 2008; Bhat et al. 2009; Pan et al. 2009; Stegger et al. 2010). Techniques for rapid detection of CC398 MRSA have been developed, such as the CC398-specific restriction–modification (R-M) test (Stegger et al. 2011); however, control therapeutics are yet to be developed. Identifying host specificity factors is the first step required in developing these targeted therapeutics, as well as in understanding the specific host–pathogen interactions of MRSA.

Molecular typing studies show that animal isolates generally belong to different lineages than human isolates. However, comparative genomics by sequencing and multi-strain microarray revealed the predicted proteins involved in binding to specific host receptors during colonization are surprisingly similar between animal and human isolates (Lowder et al. 2009; Guinane et al. 2010; McCarthy and Lindsay 2010; Sung et al. 2008; Schijffelen et al. 2010; Garcia-Álvarez et al. 2011). In contrast, some mobile genetic elements (MGEs) carrying genes encoding host immune evasion strategies appear to be host specific. The immune evasion cluster (IEC) of genes carried on the ϕ3 family of bacteriophage are found in nearly all human isolates, but in few animal isolates (Sung et al. 2008). IEC genes encode proteins that interact specifically with the human immune response; chp (chemotaxis inhibitory protein) encodes an inhibitor of chemokine responses and blocks neutrophil recruitment, sak (staphylokinase) encodes an inhibitor of defensins, and scr (staphylococcal complement inhibitor) encodes an inhibitor of complement formation and neutrophil phagocytosis (van Wamel et al. 2006). Other reported human specificity factors are the iron-regulated surface determinant B (isdB) gene that encodes a surface protein that specifically binds human hemoglobin (Pishchany et al. 2010) and Staphylococcal superantigen–like (SSL)-10 that binds to human but not animal IgG (Itoh et al. 2010; Patel et al. 2010). Recently, homologs of von Willebrand–binding protein (VwBP) and Scn proteins carried on S. aureus pathogenicity islands (SaPIs) have been described in animal genomes (Guinane et al. 2010; Schijffelen et al. 2010). The chromosomally encoded vwb (von Willebrand–binding protein) encodes a protein that interacts with von Willebrand factor (vWF) and activates prothrombin (Bjerketop et al. 2002). Viana et al. (2010) have recently shown that the SaPI-encoded vwb genes possess a variable N-terminal region that allows them to specifically coagulate ruminant plasma.

Previous comparative genomics studies of S. aureus populations using microarrays have been limited to identifying human specificity factors because the 7-strain S. aureus microarray only represents human S. aureus genomes (Witney et al. 2005; Lindsay et al. 2006; Sung et al. 2008) or use a small selection of probes (Hallin et al. 2011). Here we describe the design and validation of a new 62-strain S. aureus microarray (SAM-62), containing probes that target all the predicted open-reading frames of 62 whole-genome sequencing projects and 153 plasmid sequences; these included genomes from CC398 and other animal isolates. We compared CC398 isolates originating from pigs versus humans to identify genes or gene variants that were associated with host and included isolates from three different countries. Host-specific genes are targets in the development of diagnostics, preventatives, and therapeutics.

**Materials and Methods**

**Microarray Design**

The complete genome sequences of 62 strains of S. aureus, representing 18 different lineages and six animal isolates (2 cow, 1 sheep, 1 chicken, 2 pig) were included (Table 1). At the time of microarray design, the sequence of 40 strains was not annotated; therefore prediction of coding sequences (CDSs) was performed in AMLgene (http://www.genoscope.cns.fr/agc/tools/amlgene/). The complete sequence of 153 S. aureus plasmids was also included (supplementary table 1, Supplementary Material online). In total, the 62-strain S. aureus microarray (SAM-62) contains 29,739 60-mer oligonucleotide probes representing 6,520 genes and an additional 579 gene variants. We note that microarrays only report the presence and absence of the genes they represent.

The design strategy for the microarray was the same as previously reported for the 7-strain S. aureus microarray (Witney et al. 2005). Briefly, one strain was chosen as a base strain, and all genes were deposited in the microarray gene pool. A second strain was compared against the base strain and all genes that were strain specific or significantly divergent were added to the gene pool. The process was repeated so that the microarray gene pool was representative of all CDSs from the 62 sequenced S. aureus genomes and the 153 sequenced plasmid genomes. Multiple optimal hybridization of 60-mer oligonucleotide sequences were designed for all genes (Oxford Gene Technologies), from which a minimal nonredundant subset of oligonucleotides were selected with a target coverage of three 60-mers per gene.

Additional design was performed so that the carriage of gene variants of highly variable genes could be investigated comprehensively for the first time in S. aureus. These genes
Table 1
Sequenced Staphylococcus aureus Genomes Represented on the Microarray

| Lineage | CC | ST | Strain | Host | Infection Status | GenBank Accession Number |
|---------|----|----|--------|------|------------------|-------------------------|
| 1       | 1  | 1  | MSSA476* | H    | I                | BX571857 |
| 1       | 1  |    | MW2*     | H    | I                | BA000033 |
| 1       | 1  |    | TCH70    | H    | S                | NZ_ACHH00000000 |
| 5       | 5  | 5  | A95937   | H    | I                | NZ_ACKC00000000 |
| 5       | 5  | 5  | A6224    | H    | I                | NZ_ACKE00000000 |
| 5       | 5  | 5  | A6300    | H    | I                | NZ_ACKF00000000 |
| 5       | 5  | 5  | A8115    | H    | S                | NZ_ACKG00000000 |
| 5       | 5  | 5  | A8117    | H    | S                | NZACY01000000 |
| 5       | 5  | 5  | A9719    | H    | U                | NZ_ACKK00000000 |
| 5       | 5  | 5  | A9763    | H    | U                | NZACKH00000000 |
| 5       | 5  | 5  | A9781    | H    | U                | NZ_ACKL00000000 |
| 5       | 5  | 5  | A9299    | H    | U                | NZACKH00000000 |
| 5       | 5  | 5  | A10102   | H    | U                | NZ_ACS00000000 |
| 5       | 5  |    | CF-Marseille | H | I | NZ_CABA00000000 |
| 5       | 5  |    | ED98*   | A    | I                | CP001781 |
| 5       | 5  |    | Mu3*    | H    | I                | AP009324 |
| 5       | 5  |    | Mu50*   | H    | I                | BA000017 |
| 5       | 5  |    | N315*   | H    | S                | BA000018 |
| 5       | 5  |    | JH1*    | H    | I                | CP000736 |
| 5       | 5  |    | JH9*    | H    | I                | CP000703 |
| 7       | 7  | 7  | USA300 TCH959* | H | S | NZ_AASB00000000 |
| 8       | 8  | 8  | A5948    | H    | U                | NZ_ACKD00000000 |
| 8       | 8  | 8  | A9765    | H    | U                | NZ_ACSN00000000 |
| 8       | 8  | 8  | NCTC 8325* | H | S | CP000253 |
| 8       | 8  | 8  | Newman*  | H    | I                | AP009351 |
| 8       | 8  | 8  | USA300 FPR3757* | H | I | CP000255 |
| 8       | 8  | 8  | USA300 TCH1516* | H | S | CP000730 |
| 8       | 8  | 8  | COL*     | H    | S?               | CP000046 |
| 10      | 10 | 10 | H19      | H    | U                | NZ_ACSS01000000 |
| 10      | 10 | 10 | D139     | H    | U                | NZ_ACSR01000000 |
| 22      | 22 | 22 | EMRSA15/5096* | H | I | http://www.sanger.ac.uk |
| 30      | 30 | 30 | 55/2053  | H    | U                | NZ_ACR00000000 |
| 30      | 30 | 30 | 58-424   | H    | U                | NZ_ACUO10000000 |
| 30      | 30 | 30 | 65-1322  | H    | U                | NZ_ACUO00000000 |
| 30      | 30 | 30 | 68-397   | H    | U                | NZ_ACUO00000000 |
| 30      | 30 | 30 | A017934/97| H    | U                | NZ_ACYP01000000 |
| 30      | 30 | 30 | Btm1260  | H    | U                | NZ_ACUU01000000 |
| 30      | 30 | 30 | C101     | H    | U                | NZ_ACS01000000 |
| 30      | 30 | 30 | E1410    | H    | U                | NZ_ACUJ00000000 |
| 30      | 30 | 30 | M1015    | H    | U                | NZ_ACSS01000000 |
| 30      | 30 | 30 | M876     | H    | U                | NZ_ACV00000000 |
| 30      | 30 | 30 | M899     | H    | U                | NZ_ACSU01000000 |
| 30      | 30 | 30 | MN8      | H    | S                | NZ_ACAJ00000000 |
| 30      | 30 | 30 | TCH60    | H    | S                | NZ_ACHC00000000 |
| 30      | 30 | 30 | WBG10049 | H    | V                | NZ_ACSV01000000 |
| 30      | 30 | 30 | WW2703/97| H    | U                | NZ_ACSV01000000 |
| 34      | 34 | 34 | C160     | H    | U                | NZ_ACUV01000000 |
| 36      | 36 | 36 | MRSA252* | H    | I                | BX571856 |
| 42      | 42 | 42 | C427     | H    | U                | NZ_ACSQ01000000 |
| 45      | 45 | 45 | A9635    | H    | U                | NZ_ACKI00000000 |
| 72      | 72 | 72 | TCH130   | H    | S                | NZ_ACHD00000000 |
| 75      | 75 | 75 | MSHR 1132* | H | I | FR821777 |
| 133     | 133| 133| ED133*   | O    | I                | CP001996 |
| 151     | 151| 151| RF122/ET3-1* | B | I | A938182 |
| 152     | 152| 152| BB155*   | H    | I                | Personal communication |
| 239     | 239| 239| JKD6008  | H    | I                | NZ_ABRZ00000000 |
in *S. aureus* include the *hsdS* genes of the Saul R-M system and multiple surface and secreted genes (Waldron and Lindsay 2006; Schijffelen et al. 2010). For these genes, 60-mer oligonucleotide probes specific to variable regions were designed such that all known variants of each region of a gene were represented on the microarray.

Arrays were manufactured on the Inkjet in situ synthesized platform (Agilent Technologies) using the 8 × 60k format. The array design is available in B@G@Sbase (accession no. A-BUGS-38; http://bugs.sgul.ac.uk/A-BUGS-38) and ArrayExpress (accession no. A-BUGS-38).

### Microarray Labeling, Hybridizations, and Scanning

DNA extraction of all strains was performed using the EdgeBiosystems PureElute Bacterial Genomic kit (http://www.edgebio.com) and DNA concentration was measured using the Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies).

A total of 400 μg of test strain DNA and 100 μg of reference strain (MRSA252) of DNA were primed with random primers (2.5 μl) in a volume of 15.5 μl and heated to 95°C to separate DNA strands. Test and reference DNA samples were labeled with 40 μM Cy3-dUTP or 40 μM Cy5-dUTP, respectively, using DNA polymerase I large fragment (Exo-Klenow; Agilent Technologies) in 25 μl volumes at 37°C for 2 h before heating to 65°C for 10 min to inactivate the enzyme. Labeled DNA samples were pooled and purified using Amicon Ultra Millipore 30-kDa filters (Millipore) and eluted in Tris-ethylenediaminetetraacetic acid in a 20-μl volume.

Eighteen microliters of Cy3- and Cy5-labeled genomic DNA mixture was prepared for microarray hybridization using the Agilent Hybridization Kit (Agilent Technologies) in a total volume of 45 μl and incubated at 95°C for 3 min then 37°C for 30 min; 40 μl of the hybridization sample mixture was loaded onto an 8 × 60k microarray and hybridized for 18 h at 65°C and 20 revolutions per minute in a hybridization oven (Agilent Technologies).

After hybridization, the microarrays were washed in buffer 1–5 min at room temperature and then in buffer 2 for 1 min at 37°C (Agilent Technologies). Slides were then scanned using an Agilent High Resolution Microarray Scanner (Agilent Technologies).

### Microarray Data Analysis

Data were extracted and processed from each scanned image using Agilent Feature Extraction v10.7.3.1 (Agilent Technologies). Data from all microarrays were normalized using locally weighted scatterplot smoothing and analyzed in GeneSpring GX v11.01 (Agilent Technologies). All clustering tree analyses were performed using the hierarchical clustering algorithm and the Euclidean distance metric using defined gene lists.

Fully annotated microarray data have been deposited in B@G@Sbase (accession No: E-BUGS-120; http://bugs.sgul.ac.uk/E-BUGS-120) and also ArrayExpress (accession No: E-BUGS-120).

### Microarray Validation

The microarray was initially validated by hybridization using eight previously sequenced isolates of *S. aureus* in triplicate (MRSA252, N315, Mu50, Mu3, COL, MSSA476, MW2, S0385). Clustering analysis was performed using a method utilizing the R statistical language (Snipen et al. 2009). Briefly, distance between genomes was computed as the Manhattan distance based on predicted sequence identity for each oligonucleotide and then trees were generated using hierarchical clustering. We also validated by assessing the distribution of 60-mer oligos that represent the *hsdS* gene variants among *S. aureus* from a selection of lineages (CC1, CC5, CC8, CC9, CC22, CC30, CC398).

### Human versus Pig Specificity of CC398 Isolates

A total of 76 *S. aureus* isolates (isolated from 2003 to 2010) from the CC398 lineage were analyzed originating from three different European countries with high levels of MRSA CC398 human infections: Belgium, Denmark, and The Netherlands (Table 2). A total of 44 isolates were from pigs and 32 isolates were isolated from humans.
Results

Microarray Validation

Clustering analysis shows that microarray data from three replicates of each isolate cluster together and that isolates cluster according to clonal complex (CC) and lineage (Fig. 1A). By including at least three 60-mer probes per gene, the few 60-mers that underperformed were compensated by flanking 60-mers, and these were randomly distributed. Therefore, the data from individual microarray experiments of the same isolate are consistently reporting gene presence.

The distribution of *hsdS* gene variants between 20 *S. aureus* isolates from 7 lineages on the microarray is as expected (Fig. 1B). In other words, all *S. aureus* of the same CC lineage carry the same target recognition domain (TRD) variants, and each CC lineage has a unique combination of TRD variants. The microarray is correctly showing the presence of genes and gene variants in *S. aureus* isolates.

The Core Genomes of CC398 Isolates Do Not Vary

The 76 CC398 isolates that we analyzed using SAM-62 did not cluster by host or geographic origin (data not shown). There was no dominant human invasive or human colonization subgroup or dominant human subgroup or pig subgroup of CC398.

All CC398 isolates essentially carried the same core genome–encoded surface genes, secreted genes, and their variants (data not shown). Absence of a surface or secreted gene in an isolate was observed but rare, for *fnbpB* (*n* = 1/76 isolates), *srdD* (*n* = 1/76 isolates), and *sdrE* (*n* = 6/76 isolates). There was no difference in gene carriage between human invasive and colonization isolates, between pig and human isolates, or between isolates from different countries. These findings agree with whole-genome sequence analysis studies showing that carriage and variation in these genes is determined by lineage (Schijffelen et al. 2010).

Distribution of MGEs

The distribution of bacteriophage, SaPIs and plasmids that carry virulence genes, resistance genes and genes with putative roles in host specificity was highly variable (Figs. 2 and 3). This suggests that MGEs frequently move among CC398 *S. aureus*. CC398 *S. aureus* commonly carried ϕ2 and ϕ6 bacteriophage and SaPI5, whereas other bacteriophage and SaPIs were less frequent or absent. We found substantial mosaicism in all bacteriophage and SaPIs, as expected (Lindsay and Holden 2006); 14/18 different plasmid *rep* genes were positive in at least one CC398 isolate, suggesting that there is great diversity in plasmids circulating among the CC398 lineage. Interestingly, genes encoding for conjugation transfer, the *tra* genes, were not present in any of the CC398 isolates that we analyzed.

The distribution of transposons and integrative and conjugative element (ICE) 1 and ICE2 (Schijffelen et al. 2010) were less variable among the CC398 isolates. Transposons found in the sequenced CC398 isolate (including Tn552, Tn916) were widely distributed among the CC398 isolates we analyzed (data not shown) (McCarthy and Lindsay 2010). Carriage of both ICE1 (3/76 isolates) and ICE2 (1/76 isolates) of the sequenced isolate was rare. We note that no significant differences in the distributions of MGEs between MRSA and MSSA isolates were observed.

Distribution of MGEs and Host Specificity Genes between Human versus Pig Isolates

The ϕ3 bacteriophage and the IEC genes were carried in human CC398 isolates and not in pig CC398 isolates (χ² = 6.875, *P* = 0.01) (Fig. 2). This supports evidence that these

| Table 2  | CC398 *Staphylococcus aureus* Isolates Analyzed by Microarray |
|----------|---------------------------------------------------------------|
| Host     | Country                  | Number MRSA of isolates | Number MSSA of isolates |
| Human    | Colonization             | Belgium                 | 2                        | 0                        |
|          |                          | Denmark                 | 7                        | 2                        |
|          |                          | The Netherlands         | 3                        | 1                        |
|          | Invasive                 | Belgium                 | 4                        | 3                        |
|          |                          | Denmark                 | 3                        | 0                        |
|          |                          | The Netherlands         | 7                        | 0                        |
| Pig      | Colonization             | Belgium                 | 26                       | 6                        |
|          |                          | Denmark                 | 19                       | 3                        |
|          |                          | The Netherlands         | 13                       | 0                        |
|          | Invasive                 | Belgium                 | 0                        | 0                        |
|          |                          | Denmark                 | 1                        | 2                        |
|          |                          | The Netherlands         | 0                        | 0                        |
|          | Total                    |                         | 39                       | 5                        |

**Note.**—CC398 isolates analyzed by SAM-62 originate from three countries with the highest prevalence of CC398 in Europe.
genes are associated with human specificity (van Wamel et al. 2006; Sung et al. 2008).

However, carriage was only present in a small number (5/17) of human-invasive isolates, indicating that they are not essential for human adaptation of CC398. Furthermore, all the human isolates were from invasive infection (5/17), whereas none of the human colonization isolates carried the u3 phage (0/15). This not only shows an association of the u3 with invasive infection (v2 4.411, P 0.03), but also shows the bacteriophage is not essential for infection. No other MGE was found to be associated with human invasive versus colonization isolates.

Apart from the u3 genes, several host-specific genes have recently been proposed, but our data did not confirm their association with pig specificity. First, the isdB gene has recently been shown to bind to human but not mouse hemoglobin (Pishchany et al. 2010). The variant used in the binding studies was from strain Newman, which is CC8.

We have recently shown that there is genetic variation in the isdB gene, including in the proposed host–protein-binding domains, and variants are found in various human lineages (McCarthy and Linsday 2010). Whether all the variants can bind to human hemoglobin is currently unknown. The CC398 isolate has an isdB gene that closely matches that found in the human Staphylococcus aureus lineages of CC1, CC30, and CC239. The microarray can differentiate between three types of isdB gene. In this study, all the human and pig isolates carried the same variant. It seems unlikely that this variant can bind to both human and pig hemoglobin. This suggests that the isdB gene may not play an essential role in host adaptation between humans versus pigs.

Second, the ssl-10 encodes a protein reported to bind specifically to human IgG (Itoh et al. 2010; Patel et al. 2010). SSL-10 is part of a large family of related ssl genes, which vary substantially between lineage, and the binding
ligands of most variants are unknown. However, in this study, we found no difference between the ssl-10 genes in human and pig isolates. Therefore, the ability of SSL-10 to bind to human IgG may be important, but this gene does not appear to be lost in pig isolates.

The third family of genes are the fll and flr genes encoding FLIPr and FLIPr like. There is substantial variation in these genes that is associated with S. aureus lineage. The genes are reported to bind specifically to formyl peptide receptor like 1 (FPRL1) on the leukocyte cell surface–inhibiting leukocyte responses (Prat et al. 2006; Jongerius et al. 2007). However, in this study, there was no variation between these genes in the pig versus human isolates, again arguing against an essential role for these genes in host adaptation.

SaPI5 has been shown to carry homologs of vwb and scn genes proposed to have roles in ruminant specificity and, therefore, are expected to be more prevalent in pig isolates (Guinane et al. 2010; Schijffelen et al. 2010). However, our results did not show this; instead, in all three countries, SaPI5 was not associated with pig isolates (Belgium $\chi^2 = 1.66, P = 0.20$; Denmark $\chi^2 = 2.45, P = 0.12$; Netherlands $\chi^2 = 0.34, P = 0.56$). Each isolate that possessed SaPI5 int also possessed the vwb and scn homologs (Fig. 2). Collectively, these data argue that SaPI5 and the virulence genes carried are not essential for CC398 pig specificity.

**Distribution of MGEs between Countries**

We found that some MGEs were more/less frequent in one country than others (Figs. 2 and 3). $\phi 6$ bacteriophage was more frequent in Denmark, whereas SaPI5 were less frequent in Denmark; rep$_{10}$ was less frequent in Belgium, and rep$_{26}$ was more frequent in The Netherlands. Interestingly, we found some MGEs ($\phi 2$, $\phi 3$, $\phi 6$, SaPI5, and plasmids), virulence genes and resistance genes are more frequent in one host than another, but for specific countries (Figs. 2 and 3). $\phi 2$ bacteriophages are present in more human isolates than pig isolates in Denmark, but not in Belgium or The Netherlands. $\phi 6$ bacteriophages are present.
in more pig isolates than human isolates in Belgium only. SaPI5 is more frequent in different hosts in different countries; in Belgium, carriage is most frequent in pig isolates, but in Denmark and The Netherlands, carriage is most common in human isolates. The distribution of some rep genes is also more frequent in either pig or human isolates; but trends were country specific (Fig. 3).

![Fig. 3.—Distribution of plasmid rep genes in CC398 isolates. Isolates are represented by vertical lines, and information about the host and country (BE = Belgium, DK = Denmark, NL = The Netherlands) origin of each isolate is shown at the bottom of the figure. Human isolates are colored light blue (colonization) and dark blue (invasive), and pig isolates are colored light pink (colonization) and dark pink (invasive). Horizontal lines represent 67 different 60-mer oligo probes specific to 18 plasmid rep gene families. The color in the middle depicts whether the gene is present in the respective isolate; red/yellow = present, black/blue = absent. The intensity of all the colors is an indicator of the total signal intensity, whereas the color is an indicator of test signal over reference signal ratio.](image)

### Distribution of Resistance and Virulence Genes

Twenty-seven resistance and 20 virulence genes carried on MGEs are represented on SAM-62 (Fig. 4). Resistance genes such as *blaZ* (97%), *cadDX* (33%), *dfrG* (75%), *ermC* (30%), *meca* (86%), *tetM* (96%), and *tetK* (61%) were highly prevalent or common among the CC398 isolates. Other resistance genes were also detected but at lower frequencies: *ermA* (4%), *ermB* (8%), *aac/aph* (11%), *aadD* (11%), *dfrA* (3%), *smr* (1%), and *ant4*’ (16%). The vanA gene encoding glycopeptide resistance was not detected.

The distribution of some resistance genes was host specific and geographic specific; for example, the *aac/aph* genes were associated with isolates from Belgium and The Netherlands but not Denmark ($\chi^2 = 8.78, P = 0.01$), whereas the *dfrG* was most frequent in pig isolates in Belgium, but human isolates in The Netherlands (Fig. 4). We found that the virulence genes *lukAB* and exfoliative toxin A (*eta*) were present in all CC398 isolates (data not shown). We found PVL genes (*lukFS*, three isolates) and staphylococcal enterotoxin B (*seb*, one isolate) in human isolates from Belgium and Denmark. Both these genes appear on MGEs and have the potential to disseminate through the lineage.

### Discussion

This study shows that human and animal CC398 *S. aureus* isolates are surprisingly similar, do not form human and pig subgroups and that no single gene is essential for human or pig specificity across all geographic regions. In contrast,
we found the MGE content of CC398 was highly variable. This included MGEs carrying genes for proteins that are only functional in certain mammalian hosts. These are likely candidates for rapid gene loss or acquisition as a strain adapts to a new mammalian host. However, we found no evidence that any of these genes were essential for colonization or infection of human or pig hosts. This is unexpected because each of these proteins is predicted to interact specifically with a host protein that varies substantially between mammalian hosts. This supports our previous proposal that _S. aureus_ encodes multiple surface proteins that interact with a wide range of ligands from multiple host species (McCarthy and Lindsay 2010) and that no individual protein is essential.

The one exception was the cluster of genes for human-specific immune evasion encoded on the φ3 bacteriophage. Although they were not found in all human isolates, they were notably found only in human-invasive isolates, suggesting that they may indeed play a useful if not essential role in invasive disease. φ3 was missing in all pig isolates, confirming it is not required in pigs.

We found that carriage of MGEs among the CC398 lineage is highly variable arguing that horizontal gene transfer (HGT) of MGEs in CC398 is frequent. HGT leads to the emergence of more virulent and multiresistant MRSA strains (Waldron and Lindsay 2006; Sung and Lindsay 2007). The absence of _tra_ genes for plasmid conjugation in CC398 suggests that HGT is predominantly due to transduction (Lindsay and Holden 2006). We have previously proposed that certain MGEs are more prevalent in some lineages due to control of HGT by the Sau1 R-M system (Waldron and Lindsay 2006). This is supported by recent findings of Guinane et al. (2010) showing that SaPIs are associated with ruminant _S. aureus_ isolates of lineage CC133 but not ruminant isolates of other lineages (CC97, CC126, CC151).

Carriage of certain MGEs in CC398 was most frequent in one host than another, but intriguingly, these patterns were country specific. Geography-specific differences in host distributions were seen for bacteriophage, SaPIs, and plasmids. Although all three countries share some similar MGEs, the frequency of carriage was remarkably different. The data suggest that HGT of MGEs is frequent among CC398 and transfer occurs at highest frequency between populations of bacteria in close contact. We would therefore predict that CC398 populations in other European countries or in North America will show similar trends.
America would carry their own unique MGEs at their own frequency.

This variation extended to virulence and resistance genes carried on MGE. PVL was detected in human isolates, but carriage was rare (2/9 from Belgium, 1/12 from Denmark, 0/11 from The Netherlands) and country specific despite the wide distribution of ϕ2 bacteriophage in the CC398 lineage. Likewise, resistance genes carried on plasmids are distributed differentily between host populations, but trends were country specific. In contrast, genes that encode resistance to erythromycin, tetracycline, and trimethoprim were very common in all the CC398 isolates (de Neeling et al. 2007), suggesting their importance for maintaining populations of CC398.

MGEs play a central role in the adaptation to different environments and evolution of S. aureus. Here, we have shown that carriage of MGEs is highly variable and differs between populations within the same S. aureus lineage. This study therefore argues that HGT of MGEs within lineage CC398 is frequent and that exchange is greatest among populations in close geographical contact. The presence of different MGE profiles in different populations could be a feature that is exploited for determining the epidemiological association of isolates from the same lineage. It also argues that for studies comparing strain populations for features associated with host or pathogenesis, it is essential that appropriate control populations from the same locations and time period must be collected. Our results highlight the need to better understand MGE profiles of populations and MGE transmission dynamics in order to help develop preventative strategies that are likely to reduce the incidence of infection.

**Supplementary Material**

Supplementary table 1 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

We thank Ad C. Fluit and collaborators for early provision of the whole-genome sequence of a CC398 (S0385) isolate. We thank Henrik Hasman and collaborators for the early provision of two CC398 isolates (98 and 351) for validation of the array. We thank J. Ross Fitzgerald and collaborators for early provisions of whole-genome sequences of an avian CC5 (ED98) isolate and ovine CC133 (ED133) isolate. We thank Matthew T. G. Holden and collaborators for the early provision of the whole-genome sequences of a bovine CC425 (LGA251) isolate, a human CC75 isolate (MSHR 1132), and a human CC152 (BB155) isolate. This work was supported by the PILGRIM grant from the EU FP7, and The Bacterial Microarray Group at St George’s (BpG@S; http://www.bugs.sgul.ac.uk) are funded by The Wellcome Trust.

**Literature Cited**

Bhat M, et al. 2009. Staphylococcus aureus ST398, New York City and Dominican Republic. Emerg Infect Dis. 15:285–287.
Bjerketorp J, et al. 2002. A novel von Willebrand factor binding protein expressed by Staphylococcus aureus. Microbiology 148:2037–2044.
de Neeling AJ, et al. 2007. High prevalence of methicillin resistant Staphylococcus aureus in pigs. Vet Microbiol. 122:366–372.
Denis O, et al. 2009. Methicillin-resistant Staphylococcus aureus ST398 in swine farm personnel. Belgium Emerg Infect Dis. 15:1098–1101.
García-Alvarez L, et al. 2011. Methicillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. Lancet Infect Dis. 11:595–603.
Guinane CM, et al. 2010. Evolutionary genomics of Staphylococcus aureus reveals insights into the origin and molecular basis of ruminant host adaptation. Genome Biol Evol. 12:454–466.
Hallin M, et al. 2011. Diversity of accessory genome of human and livestock-associated ST398 methicillin resistant Staphylococcus aureus strains. Infect Genet Evol. 11:290–299.
Jongerius I, et al. 2007. Staphylococcal complement evasion by various convertase-blocking molecules. J Exp Med. 204:2461–2471.
Khanna T, Friendship R, Dewey C, Weese JS. 2008. Methicillin resistant Staphylococcus aureus colonization in pigs and pig farmers. Vet Microbiol. 30:298–303.
Lewis HC, et al. 2008. Pigs as source of methicillin-resistant Staphylococcus aureus CC398 infections in humans. Denmark Emerg Infect Dis. 14:1383–1389.
Lindsay JA, Holden MT. 2006. Understanding the rise of the superbug: investigation of the evolution and genomic variation of Staphylococcus aureus. Funct Integr Genomics. 6:186–201.
Lindsay JA, et al. 2006. Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. J Bacteriol. 188:669–676.
Loeffler A, et al. 2005. Prevalence of methicillin-resistant Staphylococcus aureus among staff and pets in a small animal referral hospital in the UK. J Antimicrob Chemother. 56:692–697.
Lowder BV, et al. 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. Proc Natl Acad Sci USA. 106:19545–19550.
McCarthy AJ, Lindsay JA. 2010. Genetic variation in Staphylococcus aureus surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. BMC Microbiol. 10:173.
Pan A, et al. 2009. Community-acquired methicillin-resistant Staphylococcus aureus ST398 infection, Italy. Emerg Infect Dis. 15:845–847.
Patel D, Wines BD, Langley RJ, Fraser JD. 2010. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. J Immunol. 1:6283–6292.
Pishchany G, et al. 2010. Specificity for human hemoglobin enhances Staphylococcus aureus infection. Cell Host Microbe. 8:544–550.
Prat C, Bestebroer J, de Haas CJ, van Striph JA, van Kessel KP. 2006. A new staphylococcal anti-inflammatory protein that antagonizes the formyl peptide receptor-like 1. J Immunol. 177:8017–8026.
Schijffelen MJ, Boel CH, van Striph JA, Fluit AC. 2010. Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis. BMC Genomics. 11:376.
Smith TC, et al. 2009. Methicillin-resistant Staphylococcus aureus (MRSA) strain ST398 is present in midwestern U.S. swine and swine workers. PLoS One. 4:e4258.

Snipen L, Nyquist OL, Solheim M, Aakre A, Nes IF. 2009. Improved analysis of bacterial CGH data beyond the log-ratio paradigm. BMC Bioinformatics. 19:91.

Stegger M, Lindsay JA, Moodley A, Skov R, Broens EM. 2011. Rapid PCR detection of Staphylococcus aureus clonal complex 398 by targeting the restriction-modification system carrying sau1-hsdS1. J Clin Microbiol. 49:732–734.

Stegger M, Lindsay JA, Sørum M, Gould KA, Skov R. 2010. Genetic diversity in CC398 methicillin-resistant Staphylococcus aureus isolates of different geographical origin. Clin Microbiol Infect. 16:1017–1019.

Sung JM, Lindsay JA. 2007. Staphylococcus aureus strains that are hypersusceptible to resistance gene transfer from enterococci. Antimicrob Agents Chemother. 51:2189–2191.

Sung JM, Lloyd DH, Lindsay JA. 2008. Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology 154:1949–1959.

van Belkum A, et al. 2008. Methicillin-resistant and -susceptible Staphylococcus aureus sequence type 398 in pigs and humans. Emerg Infect Dis. 14:479–483.

van Cleef BA, et al. 2011. Livestock-associated methicillin-resistant Staphylococcus aureus in humans, Europe. Emerg Infect Dis. 17:502–505.

van Wamel WJ, Rooijakkers SH, Ruyken M, van Kessel KP, van Strijp JA. 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages. J Bacteriol. 188:1310–1315.

Viana D, et al. 2010. Adaptation of Staphylococcus aureus to ruminant and equine hosts involves SaP carried variants of von Willebrand factor-binding protein. Mol Microbiol. 77:1583–1594.

Waldron DE, Lindsay JA. 2006. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into Staphylococcus aureus and between S. aureus isolates of different lineages. J Bacteriol. 188:5578–5585.

Welinder-Olsson C, et al. 2008. Infection with Panton-Valentine leukocidin-positive methicillin-resistant Staphylococcus aureus t034. Emerg Infect Dis. 14:1271–1272.

Witney AA, et al. 2005. Design, validation, and application of a seven-strain Staphylococcus aureus PCR product microarray for comparative genomics. Appl Environ Microbiol. 71:7504–7514.

Witte W, Strommenger B, Stanek C, Cuny C. 2007. Methicillin-resistant Staphylococcus aureus ST398 in humans and animals, Central Europe. Emerg Infect Dis. 13:255–258.

Wulf M, Voss A. 2008. MRSA in livestock animals—an epidemic waiting to happen? Clin Microbiol Infect. 14:519–21.

Wulf MW, et al. 2008. First outbreak of methicillin-resistant Staphylococcus aureus ST398 in a Dutch hospital, June 2007. Euro Surveill. 13:8051.

Associate editor: George Weinstock