Evolutionary Genomics of *Staphylococcus aureus* Reveals Insights into the Origin and Molecular Basis of Ruminant Host Adaptation

Caitriona M. Guinane[1], Nouri L. Ben Zakour[1], Maria A. Tormo-Mas[2], Lucy A. Weinert[3], Bethan V. Lowder[1], Robyn A. Cartwright[1], Davida S. Smyth[4], Cyril J. Smyth[5], Jodi A. Lindsay[6], Katherine A. Gould[6], Adam Witney[6], Jason Hinds[6], Jonathan P. Bollback[3], Andrew Rambaut[3], José R. Penadés[2], and J. Ross Fitzgerald*[,1]

[1]The Roslin Institute and Centre for Infectious Diseases, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Scotland, United Kingdom

[2]Centro de Investigación y Tecnología Animal, Instituto Valenciano de Investigaciones Agrarias, Castellón, Spain

[3]Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh, Scotland, United Kingdom

[4]University of Mississippi Medical Center, Jackson, Mississippi

[5]Department of Microbiology, School of Genetics and Microbiology, Moyne Institute of Preventive Medicine, University of Dublin, Trinity College, Dublin, Ireland

[6]Department of Cellular & Molecular Medicine, St George's Hospital Medical School, University of London, London, United Kingdom

*Corresponding author: E-mail: ross.fitzgerald@ed.ac.uk.

ED133 genome sequence: accession no CP001996.

Accepted: 4 June 2010

Abstract

Phenotypic biotyping has traditionally been used to differentiate bacteria occupying distinct ecological niches such as host species. For example, the capacity of *Staphylococcus aureus* from sheep to coagulate ruminant plasma, reported over 60 years ago, led to the description of small ruminant and bovine *S. aureus* ecovars. The great majority of small ruminant isolates are represented by a single, widespread clonal complex (CC133) of *S. aureus*, but its evolutionary origin and the molecular basis for its host tropism remain unknown. Here, we provide evidence that the CC133 clone evolved as the result of a human to ruminant host jump followed by adaptive genome diversification. Comparative whole-genome sequencing revealed molecular evidence for host adaptation including gene decay and diversification of proteins involved in host–pathogen interactions. Importantly, several novel mobile genetic elements encoding virulence proteins with attenuated or enhanced activity in ruminants were widely distributed in CC133 isolates, suggesting a key role in its host-specific interactions. To investigate this further, we examined the activity of a novel staphylococcal pathogenicity island (SaPIov2) found in the great majority of CC133 isolates which encodes a variant of the chromosomally encoded von Willebrand-binding protein ($vWbp^{Sov2}$), previously demonstrated to have coagulase activity for human plasma. Remarkably, we discovered that SaPIov2 confers the ability to coagulate ruminant plasma suggesting an important role in ruminant disease pathogenesis and revealing the origin of a defining phenotype of the classical *S. aureus* biotyping scheme. Taken together, these data provide broad new insights into the origin and molecular basis of *S. aureus* ruminant host specificity.

Key words: bacteria, mobile genetic elements, genome diversification, population genetics, niche adaptation.

Introduction

Ancient domestication and the recent globalization of the livestock industry have resulted in increased opportunities for the transfer of bacteria between human and animal hosts and their subsequent dissemination. Several studies have revealed the capacity of bacterial pathogens to switch and adapt to different host species leading to host restriction but the molecular basis remains poorly understood (Eppinger et al. 2006; Herron-Olson et al. 2007; Lefebure and Stanhope 2007; Lowder et al. 2009).

Staphylococcus aureus is a major human and animal pathogen which is responsible for a large proportion of ruminant mastitis infections worldwide. In the 1930s, phenotypes unique to animal strains of *S. aureus* were first observed and subsequently used to define specific ecological variants (ecovars) associated with different host species (Madison 1935; Minnett 1936; Devriese 1984). For example, a feature of the small ruminant and bovine ecovars which differentiated them from human or poultry strains was the ability to coagulate plasma from ruminants, but the molecular basis for this phenotype remains unknown. Other determinants which comprised the biotyping scheme included production of β-hemolysin, Staphylokinase, and crystal violet growth reaction (Devriese 1984). These phenotypic data suggested that different *S. aureus* strains have evolved unique traits that are dependent on their host habitat. Subsequently, numerous population genetic studies have identified the existence of genotypes of *S. aureus* that are associated with cows, sheep, and goats but rarely isolated from humans, suggesting that they are specialized for ruminant hosts (Kapur et al. 1995; Fitzgerald et al. 1997; Jorgensen et al. 2005; Smyth et al. 2009). Of note, several studies have reported the existence of a single clonal complex (CC133) identified by multilocus sequence typing (MLST) which is responsible for the majority of intramammary infections of small ruminants including sheep and goats and can also cause mastitis in cows (Jorgensen et al. 2005; Aires-de-Sousa et al. 2007; Ben Zakour et al. 2008; Smyth et al. 2009). However, the evolutionary origin and the molecular basis for its unique broad ruminant host tropism is unknown. Previously, the genome sequence of a bovine-specialized strain of *S. aureus* (strain RF122; ST151) provided evidence that bovine *S. aureus* had diversified from an ancestor that resembles strains of human origin through a combination of the acquisition of mobile genetic elements (MGEs) and gene decay (Herron et al. 2002; Herron-Olson et al. 2007). Here, we employ a combination of population genetics, comparative genomics, and ex vivo functional analysis to investigate the origin and genetic basis for the small ruminant host tropism of the *S. aureus* CC133 lineage, resulting in broad new insights into bacterial adaptation to livestock animals.

Materials and Methods

Bacterial Strains, Growth Conditions, and Genomic DNA Isolation

A total of 29 *S. aureus* isolates from intramammary infections of cows, sheep, and goats in nine countries on three different continents were employed (supplementary table S1, Supplementary Material online). Strains were grown in Brain Heart Infusion Broth or Tryptic Soy Broth (TSB) at 37 °C with shaking at 200 rpm. DNA was extracted from 1-ml volumes of overnight TSB cultures using the Edge Biosystems Bacterial Genomic DNA Purification Kit (Edge Biosystems) according to the manufacturers instructions with the addition of lysostaphin (AMBI Products LLC) (5 mg/ml) to the cell lysis step.

Genome Sequencing, Assembly, and Bioinformatic Analysis

Massively parallel 454 pyrosequencing of genomic DNA from ovine strain ED133 to a coverage of 36.5× was carried out by 454 Life Sciences (www.454.com) with a GS20 sequencer followed by assembly into 78 contigs using the Newbler program (Roche). Order and orientation of assembled contigs were determined by scaffolding with the published genome sequence of the bovine strain RF122 (Herron-Olson et al. 2007). Primers specific for gap edges were designed using Projector 2 software (van Hijum et al. 2005) for polymerase chain reaction (PCR) amplification of gap regions using Platinum Hi-fidelity PCR Supermix (Invitrogen) or *Pfu* DNA polymerase (Promega) with a Biometra TGradient thermocyler followed by directed sequencing of PCR products by primer walking. Whole-genome assembly was performed using the PHRED–PHRAP–CONSED package (Ewing et al. 1998; Gordon 2003). The contigs specific for prophage regions of the genome were closed by combinatorial PCR followed by cloning of products into the pSC-A-amp/kan or pSC-B-amp/kan vectors using the StrataClone PCR cloning kit or Blunt PCR cloning kits, respectively (Agilent Technologies), and subsequently sequenced by primer walking. The structure of each phage was verified by overlapping PCR, followed by restriction digestion with a range of restriction endonucleases including *Hin*P1, *Alu*I, *Eco*R1, and *Hin*dIII (New England Biolabs) (data not shown). The complete whole-genome sequence was verified by pulsed field gel electrophoresis of ED133 after genomic DNA restriction digestion with endonuclease *Sma*I (data not shown). A Perl script was written to determine the base quality of the genome sequence, and 99.95% of the genome was Q40+. Because of the existence of multiple repeat regions and extensive interphage homology (fig. 2), the quality score for each phage was determined separately (ΦSaov1 330700 nt–335057 nt; 99.6% Q40, ΦSaov2 1116584 nt–1122000 nt; 99.4% Q40, ΦSaov3 2002310 nt–2008800 nt; 99.4% Q40).

For genes exhibiting >85% identity with orthologous sequences from other *S. aureus* genomes, annotation was duplicated from bovine strain RF122 and the human strain COL using GATU software (Tcherepanov et al. 2006), and the remaining genes were annotated individually. The genome sequence was also submitted to the J. Craig Venter Institute (www.jcvi.org) for automated pipeline annotation using Glimmer, Blast_Extend_Repraze (BER) alignments, alignments with experimentally characterized genes, hidden Markov model matching, and searches for biologically significant patterns using PROSITE, and the data were manually

curated in Artemis (Rutherford et al. 2000; Carver et al. 2008). Complementary annotation data were provided by the SEED (Overbeek et al. 2005) and the RAST annotation servers (Aziz et al. 2008), and genome comparisons were made using the Artemis comparison tool (Carver et al. 2005).

Pseudogene Identification

Open reading frames (ORFs) displaying evidence of a frameshift were indicated by the BER search performed in the JCVI annotation pipeline and manually checked. To complement this, a Perl script was written for the identification of truncated ORFs in ED133, caused by mutations leading to a premature stop codon, or indels causing a frameshift, wherein the sequence of each gene was compared with its closest homolog as determined by the BER search. Size differences, as a percentage of the gene length and percentage identity to best hit, were reported and regions manually checked to eliminate differences caused by variant start site predictions. Mutations were confirmed by PCR amplification and conventional Sanger sequencing. Pseudogenes were not determined for integrated prophage due to lower base quality scores.

Phylogenetic Analysis

Reconstruction of evolutionary relationships was carried out using the MEGA 4 package (Tamura et al. 2007). Concatentated MLST sequence data obtained from the MLST online database (http://saureus.mlst.net) was used to construct a consensus Neighbor-Joining tree from 500 bootstrapping replicates.

In order to date the predicted host switch from humans to ruminants within the CC133, a dated phylogeny of all members of the complex was estimated using the phylogenetics program BEAST (Drummond and Rambaut 2007) was used, employing its uncorrelated lognormal relaxed clock model, in which rates of evolution are allowed to vary across all branches (Drummond et al. 2006) and used the Hasegawa–Kishino–Yano model of nucleotide substitution (Hasegawa et al. 1985) assuming a gamma distribution of substitution rate across sites in the alignment. Six separate chains were run for 10,000,000 steps each, and after discarding the first 10% as burn-in, convergence was verified using the program Tracer in the BEAST package. The tree was calibrated with a mutation rate (3.3×10^{-6} per site per year) estimated by Harris et al. (2010) for the ST239 clone using dated samples over a 21-year period. This rate is consistent with the rate calculated by Smyth et al. (2010) for the same clone ($3.3–4.6 \times 10^{-6}$ per site per year) and Lowder et al. (2009) who used dated samples over a 30-year period from a different clonal complex (ST5) (5.125×10^{-6} per site per year). Default prior probability distributions were used for all other parameters. The date of the host jump must have occurred between the date of the most common recent ancestor of the Bovidae-infecting clade and the node ancestral to that (i.e, the common ancestor of the Bovidae clade and ST945). We therefore estimated the date of the switch as the lower bound of the 95% highest posterior density (HPD) from the most common recent ancestor of the Bovid clade and the upper bound of the 95% HPD of the node ancestral to this, as inferred from the maximum posterior consensus tree. The date of the switch was predicted as the median of the estimates of the dates of these two nodes, as inferred from the maximum posterior consensus tree.

Comparative Genomic Hybridizations

A total of 2 μg of test and reference strain genomic DNA was labeled with Cy3 or Cy5 dye with DNA polymerase I large fragment (Klenow; Invitrogen), pooled, and hybridized to an *S. aureus* microarray overnight as described (Lindsay et al. 2006). A previously constructed *S. aureus* microarray, representing seven different *S. aureus* strains of human origin (Lindsay et al. 2006), was supplemented with sequences unique to the strain RF122 genome and 57 sequences specific for strain ED133. Microarrays were scanned using a Genepix Personal 4100 scanner (GRI), and data analysis was performed using Bluefuse for Microarrays 2.0 (BlueGnome) and GeneSpring 6.2 (Silicon Genetics) as described (Lindsay et al. 2006). Presence/absence calls were made based on a 2-fold cut off. Selected representative genes predicted to be present or absent by comparative genome hybridization (CGH) analysis were confirmed by gene-specific PCR (data not shown). For MGE, PCR of at least two specific regions within the element were used to confirm the microarray results.

Genome-Wide Selection Analysis

ORF sequences from strains ED133 (ovine ST133), RF122 (bovine ST151), and representative human strains MRSA252 (human ST36), MSSA476 (human ST1), and USA300 (human ST8), were used to format a Blast database using the National Center for Biotechnology Information formatdb tool, and groups of orthologous genes were identified from this database using a reciprocal Blast Python script (Petersen et al. 2007). Briefly, Blast analysis of all genes from *S. aureus* strain ED133 was carried out with the database of concatenated genomes (*e*-value cut off of 0.0001). Genes were considered orthologous based on a positive match in each of the four other genomes and on reciprocal best Blast with the ED133 genome. Sequences were translated and aligned using ClustalW (Thompson et al. 1994), and phylogenetic trees were inferred for each ortholog group using the PAUP* software package (Wilgenbusch and Swofford 2003). Genes under positive selection were determined using PAML (Yang 1997). Positive diversifying selection was determined by comparing the M1a with M2a

models and the M7 with M8 (Yang 1997; Yang and Nielsen 2000) models. Genes for which the M2a and M8 models significantly better were determined to be under positive selection. Because of the large number of tests performed statistical was determined using the false detection rate using the fdrtool R package (Strimmer 2008). Individual codons of ED133 genes under positive selection were identified using the Bayes Empirical Bayes test (Yang et al. 2005) implemented in the PAML package. Recombination can give rise to false signals of positive selection based on d*N*/d*S* ratios because the methods used assume a common phylogeny for all sites (Anisimova et al. 2003; Shriner et al. 2003). Therefore, we tested genes showing evidence of positive selection for evidence of recombination among strains, using the single breakpoint analysis and KH test as implemented in HyPhy (Kosakovsky Pond et al. 2005). In brief, the method compares a likelihood model assuming a single recombination breakpoint with two different topologies on each side of the breakpoint, with a model that assumes no recombination. If support for a model of recombination was found the KH test (Kishino and Hasegawa 1989) for incongruence, as implemented in HyPhy, was used to determine if it was significant.

Genetic Manipulation of *S. aureus* Strains

Staphylococcus aureus strains and plasmids used for functional analyses are outlined in supplementary table S2 (Supplementary Material online). A SaPIov2 derivative with *tetM* inserted into a noncoding region of the island was constructed by allele replacement with a plasmid constructed by cloning the *tetM* gene flanked by SaPIov2-specific sequences into plasmid pRN6680 (Sloane et al. 1991). The SaPIov2 sequences were amplified using oligonucleotides SaPIov2-1mSXb (GATTACAAAACTAAAATCTGAC) and SaPIov2-2cB (CTCATTATTACTTTATTGACC), and SaPIov2-3mP (ACTTTAATTAAAAAACATCACTTC) and SaPIov2-4cE (TAACTATATCATTTTAAACTTGC) containing 5′ restriction sites for *Eco*RI, *Pst*I, *Bam*HI, and *Xba*I, respectively. The PCR products were digested with appropriate restriction endonucleases, ligated with the *tetM* gene and cloned into plasmid pRN6680. The resulting plasmid was restriction digested at native *Eco*RI and *Hin*dIII sites and ligated into the multiple cloning site of the temperature-sensitive plasmid vector pMAD (Bruckner 1997), generating pJP766. Plasmid pJP766 was introduced by electroporation into *S. aureus* strain RN4220 before transduction into strain ED133 using phage 80α, prior to allele replacement. The temperature-sensitive phenotype of the plasmids facilitated integration by homologous recombination, and a double-crossover event was detected by plating on appropriate antibiotics followed by confirmation of a stable mutant by PCR and directed sequencing. Strain ED133 ΔSaPIov2 was obtained by plating strain ED133 SaPIov2-*tet*M on tryptic soya agar followed by replica plating onto tetracycline-containing medium to identify strain sensitive to tetracycline, and deletion of SaPIov2 was confirmed by PCR and Southern blot analysis. Strain RN4220 SaPIov2 was generated by transduction of SaPIov2-*tet*M from strain ED133 SaPIov2-*tet*M to RN4220, after SOS induction of resident prophages, as previously described (Ubeda et al. 2005). Derivatives of clinical isolates VI50897, 283, VET-BZ31, and VI50896 which were deleted for SaPIov2, replaced with SaPIN1 from strain N315, were constructed by transduction of SaPIN1-*tet*M from RN4220 as previously described (Ubeda et al. 2005) followed by plating onto TSA containing tetracycline and PCR confirmation of the SaPI replacement.

Coagulase Assay

Rabbit, bovine, ovine, or caprine plasma with ethylenediaminetetraacetic acid were used for the coagulation experiments. The tube coagulation assay was performed in glass tubes by mixing 300 μl of plasma with 1×10^8 *S. aureus* bacteria from an overnight culture. The tubes were incubated at 37 °C, and the level of coagulation was observed by tilting the tubes. A positive test resulted in a coherent clot after a 4-h incubation.

Results and Discussion

Evidence for a Human Origin for the Ruminant CC133 Lineage

Numerous studies have identified the existence of host-specific genotypes of *S. aureus* (Kapur et al. 1995; Fitzgerald et al. 1997; Rodgers et al. 1999; van Leeuwen et al. 2003; Smyth et al. 2009). To examine the relatedness of ruminant-associated *S. aureus* isolates to extant human *S. aureus* genotypes within the species, we constructed a phylogenetic tree based on the concatenated MLST sequences of 130 strains selected to represent the breadth of diversity of *S. aureus* isolates of human and animal origin in the *S. aureus* MLST database (http://saureus.mlst.net) (fig. 1). The tree indicates the existence of numerous clonal complexes of closely related genotypes or lineages within the species, consistent with previously published findings regarding strains of human origin (Feil et al. 2003; Robinson et al. 2005; Cooper and Feil 2006; Lindsay et al. 2006). Of note, the majority of ruminant-associated sequence types (STs) belong to three major complexes, which include CC97, CC151, and CC133, respectively (fig. 1), indicating a narrow distribution of ruminant-associated genotypes across the species tree. Although CC97 *S. aureus* strains are commonly isolated from cows, they have also been detected among human and porcine hosts indicating a broad host tropism (Feil et al. 2003; Smith et al. 2005; Guinane et al. 2008; Battisti et al. 2009; Smyth et al. 2009). In contrast, CC151 strains are predominantly associated with cows and have not been detected among humans previously (Guinane et al. 2008). Most small ruminant (sheep

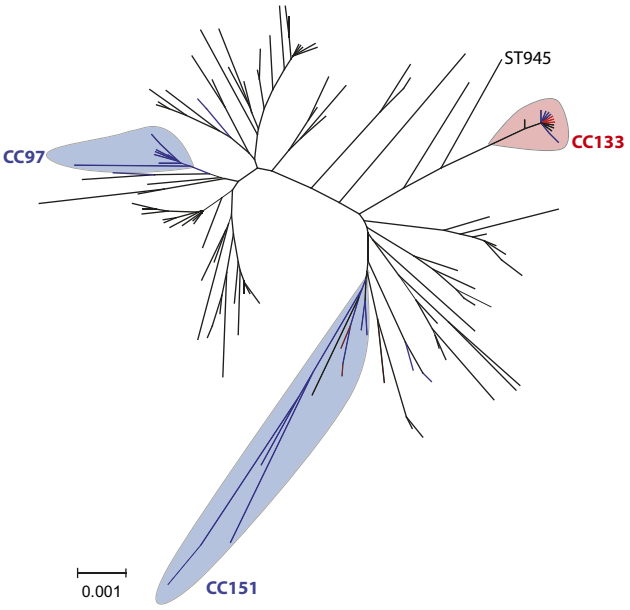

Fig. 1.—The majority of *Staphylococcus aureus* isolates from ruminants belong to three clonal lineages. Neighbor-Joining consensus tree inferred from 500 bootstrap replicates was constructed using concatenated sequences of 33 *S. aureus* STs from ruminants and 97 selected representative STs from humans and other animals. STs from large (cows) and small (sheep or goats) ruminants are depicted by blue and red branches, and the major lineages associated with large and small ruminants are indicated by blue and red shading, respectively. Branches depicting STs of human, avian, or unknown host origin are indicated in black. ST945 represents the human-associated ST, which is most closely related to CC133, used to calculate the date range for the predicted host jump.

and goat) strains belong to a single clonal complex CC133, which includes at least 23 STs, as demonstrated recently by Smyth et al. (2009), consistent with several reports of a widespread *S. aureus* clone associated with small ruminant mastitis (Jorgensen et al. 2005; Aires-de-Sousa et al. 2007; Smyth et al. 2009). Isolates of the CC133 lineage, although typically associated with sheep and goats and genetically distinct from most bovine strains, are occasionally associated with intramammary infections of cows (Jorgensen et al. 2005; Smyth et al. 2009). The broad diversity of human *S. aureus* lineages and the narrow distribution of ruminant-associated lineages in the *S. aureus* species tree imply that *S. aureus* is predominantly a human-adapted bacterium which has coevolved with its host for a longtime in evolutionary terms. In contrast, the capacity to colonize and infect ruminant host species is most likely the result of a small number of host jumps by *S. aureus* strains of human origin which were followed by genetic adaptation to ruminants, in a similar fashion to the recently reported poultry host switch by a subtype of the successful human CC5 lineage (Lowder et al. 2009). An alternative hypothesis, which can't be ruled out, is that the small number of extant ruminant-associated lineages could reflect the possibility that a limited number of *S. aureus* genotypes were already colonizing the first animals to be domesticated before undergoing clonal expansion through breeding of livestock. In contrast to the poultry CC5 clade, the existence of considerable genetic diversity in the CC133 lineage (Smyth et al. 2009) indicates that the most recent common ancestor of the CC133 ruminant-specific clone did not occur in the very recent past. In order to date the predicted host switch from humans to ruminants resulting in the CC133 lineage, we estimated a dated phylogeny of all members of the complex using the phylogenetics program BEAST (Drummond and Rambaut 2007). The tree was calibrated with a mutation rate (3.3×10^{-6} per site per year) estimated by Harris et al. (2010) for the ST239 clone using dated samples over a 21-year period. This rate is consistent with the rate calculated for the same clone ($3.3–4.6 \times 10^{-6}$ per site per year) by Smyth et al. (2010) and for the ST5 poultry clade using dated samples over a 30-year period Lowder et al. (2009) (5.125×10^{-6} per site per year). The date of the host jump must have occurred between the date of the most common recent ancestor of the ruminant-infecting clade and the node ancestral to that (i.e., the common ancestor of the Bovidae clade and the human ST945). We therefore estimated the date of the switch as the lower bound of the 95% HPD from the most common recent ancestor of the Bovid clade and the upper bound of the 95% HPD of the node ancestral to this, as inferred from the maximum posterior consensus tree. Using this approach, we estimate the host switch from humans to the family Bovidae to have occurred approximately 115–1,204 years ago. However, it is unclear whether it is appropriate to use mutation rates based on contemporary data to date this transition because comparative rates of substitution in bacteria have been shown to be lower than those based on mutation accumulation experiments (Ochman 2003). This may be due to the failure of mildly deleterious segregating mutations to reach fixation over the longer term or the rapid saturation of highly mutagenic sites. Because substitution rates based on longer timescales have not been estimated in *S. aureus*, it is impossible to know the extent to how mutation rates and substitution rates differ. Therefore, this date estimate should be taken as a minimum date of host switch with the actual date possibly being much earlier. As such, the data indicate that the host jump did not occur in the very recent past, in contrast to the recently determined host switch for *S. aureus* from humans to poultry which is predicted to have occurred about 40 years ago (Lowder et al. 2009).

Genome Sequence of a CC133 Strain (ED133) Reveals Evidence for Host-Adaptive Diversification

In order to investigate the genetic basis for the ruminant host tropism of the *S. aureus* CC133 lineage, we determined the whole-genome sequence of a representative strain

(ED133) of the predicted founder ST (ST133) isolated from an episode of ovine clinical mastitis infection in France (formerly strain 1174) (Ben Zakour et al. 2008). The 2,832,478 bp chromosome (accession number CP001996; supplementary fig. S1, Supplementary Material online) contained a total of 2,663 coding sequences of which only 26 have not been identified in the 15 *S. aureus* whole-genome sequences (www.ncbi.nlm.nih.gov) (supplementary table S3, Supplementary Material online) or among *S. aureus* phage sequences deposited in GenBank.

Several previous studies have indicated that diversification of the core and core variable components of the genome may play an important role in the adaptation of *S. aureus* to different hosts (Herron-Olson et al. 2007; Ben Zakour et al. 2008; Sung et al. 2008; Smyth et al. 2009). Comparison of the genome of *S. aureus* ED133 with human- and bovine-sequenced strains revealed considerable variation among orthologous genes encoding proteins involved in adherence, toxin production, metabolism, replication and repair, and gene regulation. Of note, cell wall-associated (CWA) proteins are involved in critical host–pathogen interactions (Clarke and Foster 2006) and as such may be under diversifying selective pressure to adapt to polymorphic receptors in different host species. In the current study, genome-wide analysis of levels of selective pressure identified by elevated ratios of nonsynonymous to synonymous substitutions indicated several CWA protein which were under diversifying selective pressure including clumping factors A and B, fibronectin-binding proteins A, and the serine aspartate-repeat proteins SdrC and SdrE (supplementary table S4, Supplementary Material online). These data indicate that surface proteins made by strain ED133 are undergoing adaptive evolution in response to their habitat, consistent with similar findings for the bovine strain RF122 (Herron-Olson et al. 2007). Of note, only one of the proteins identified to be under positive selective pressure (ClfA) contained evidence for recombination suggesting that recombination has not played a major role in the adaptive diversification of ED133. Remarkably, the CWA protein SdrD encoded by strain ED133 has undergone diversification to the extent that a 200 amino acid region (52–252) of the ligand-binding A-domain contains only 37% amino acid identity with the corresponding region of the closest homolog made by the human sequenced strain *S. aureus* JH9 (Mwangi et al. 2007). The same region in human isolates contains 98–100% identity. Allele-specific PCR indicated the presence of the novel *sdrD* allele in all CC133 isolates examined, and its absence in human and ruminant strains from different lineages with the exception of the closely related bovine lineage CC130 (data not shown) SdrD has been demonstrated to play a role in adherence of human strains of *S. aureus* to human corneocytes (Corrigan et al. 2009) and in abscess formation (Cheng et al. 2009). The diversification of the ligand-binding domain of SdrD protein suggests that it may have a unique or attenuated function in its adopted habitat.

In addition to diversification of the surface proteome of ED133, we found evidence for metabolic and regulatory gene diversification of ED133 relative to human strains. For example, the genes involved in cobalt transport are absent in ED133 in comparison with human-sequenced strains and bovine strain RF122. In common with RF122 (Herron-Olson et al. 2007), ED133 does not contain genes encoding cadmium resistance (CadD and CadX) or the ferric hydroxamate receptor. Furthermore, two genes, Saov*0148* and *hysA*, predicted to be involved in amino acid transport and polysaccharide degradation, respectively, had elevated d_N/d_S levels indicating diversifying selective pressure (supplementary table S4, Supplementary Material online). ED133 does not contain novel regulators not previously identified among human isolates but does encode the regulator PaiB in common with other ruminant strains such as RF122 which is hypothesized to influence cytolytic toxin expression levels (Guinane et al. 2008), and two putative regulators which were under diversifying selection, consistent with niche adaptation (supplementary table S4, Supplementary Material online).

Of note, several novel genetic loci were found in the ED133 genome in the functional category of DNA replication, recombination, and repair. For example, three putative helicase genes (Saov*0027*, Saov*2254*, Saov*2534*) were identified in ED133 which were not found among other *S. aureus* strains sequenced to date. Furthermore, genes encoding an insertion element, transposase, and a putative DNA helicase (Saov*0027*) were found in a 8.6-kb region of the genome of ED133 at *orfX* (the chromosomal integration site for SCC*mec* in MRSA strains) (Ito et al. 2001), but not found among other *S. aureus* genomes sequenced to date. In addition, a gene encoding a protein with homology to determinants of cell division (Saov*1727*) was also found to be under diversifying selective pressure in ED133 (supplementary table S4, Supplementary Material online). We speculate that the acquisition of novel cell division and DNA replication machinery may influence the efficiency of bacterial growth and allow coordination with host cell activities facilitating adaptation to diverse cellular environments (Wren 2000; Schoen et al. 2007). Taken together, these data indicate diversification of the surface proteome, metabolome, and replication capacity, which may in part be the result of a habitat shift to the unique environment of the ruminant udder.

Evidence for Gene Decay in the Small Ruminant Clone CC133

A common feature of bacteria undergoing niche adaptation is the loss of function of genes superfluous or

detrimental to bacteria in the new habitat (Eppinger et al. 2006; Herron-Olson et al. 2007; Stinear et al. 2007). Pseudogenes identified in the ED133 core and core variable genome belong to several functional categories including metabolism, toxins, lipoproteins, and MGE-related proteins (supplementary table S5, Supplementary Material online). Although ED133 contained fewer predicted pseudogenes than the bovine strain RF122 (Herron-Olson et al. 2007), four pseudogenes were common to both strains and were not associated with human strains sequenced to date including genes for a lipoprotein (Saov*0050*), a high affinity iron transporter (Saov*0369)*, *splA* encoding serine protease A, and a conserved hypothetical protein (Saov*0093*). Importantly, the inactivating mutation pattern for each pseudogene is distinct for each strain, indicating that they arose in parallel in each lineage, consistent with strong selective pressure for loss of function acting on these genes in the ruminant host. For example, the loss of function of the iron transporter in distinct ruminant-associated lineages is consistent with differences in the machinery required for iron acquisition by *S. aureus* in ruminant and human hosts (Herron-Olson et al. 2007). Notably, four genes encoding putative lipoproteins (supplementary table S5, Supplementary Material online) in ED133 are no longer functional, consistent with previous studies implicating a key role for lipoproteins in host–pathogen interactions and recognition by the host immune response (Bubeck Wardenburg et al. 2006). Also of note is the loss of function of genes encoding staphylococcal toxins including delta-toxin and leukotoxin E, which have important roles in human host innate immune avoidance (Raulf et al. 1990; Schmitz et al. 1997) indicating a lack of requirement for these toxins in ruminant disease pathogenesis.

The Genome of *S. aureus* Strain ED133 Contains a Unique Complement of MGEs

A major feature of the genome of ED133 is the existence of two new members of the staphylococcal pathogenicity island family (SaPIov1 and SaPIov2; fig. 2*A*), and three novel prophages (φSaov1, φSaov2, and φSaov3; fig. 2*B*) not previously identified among *S. aureus* strains sequenced to date. SaPIov1 is 14,041 bp in size and is integrated at the same chromosomal site in ED133 (adjacent to the core variable region *vsaα*), as SaPIbov1 in the genome of strain RF122 (~0.5 Mb; supplementary fig. S1, Supplementary Material online and fig. 2*A*). Furthermore, SaPIbov1 and SaPIov1 encode integrases that share 99% amino acid identity, and both encode unique host-specific variants of TSST-1, SEC, and SEL (Fitzgerald et al. 2001). The ovine-specific variants of TSST-1 and SEC encoded by SaPIov1 have previously been demonstrated to vary in biological activity in comparison with the alleles produced by bovine or human *S. aureus* strains, suggesting a host-specific functional activity (Lee et al. 1992; Deringer et al. 1997). A second novel pathogenicity island, SaPIov2, discovered in the genome of strain ED133 (fig. 2*A*) is 14,226 bp in size and is integrated at the same chromosomal site as SaPIbov3 of strain RF122 and SaPIn1 of the human strain N315 (Baba et al. 2002; Herron-Olson et al. 2007). Of note, SaPIov2 contains a gene encoding a novel von willebrand-binding protein (vWbp^{Sov2}) with 63% amino acid identity to the previously characterized vWbp encoded in the core genome of all *S. aureus* strains (Bjerketorp et al. 2002, 2004). In common with the chromosomally encoded vWbp, the SaPIov2-encoded vWbp^{Sov2} variant contains a predicted coagulase domain associated with the coagulation of mammalian plasma that is implicated in disease pathogenesis (Moreillon et al. 1995). In addition, adjacent to the *vwb*^{Sov2} gene in SaPIov2 is a gene specific for a protein with 53% identity to the previously characterized staphylococcal complement inhibitor (SCIN) encoded by β-hemolysin converting phages of human strain origin (Rooijakkers et al. 2006, 2007). SCIN contributes to immune avoidance by inhibiting phagocytosis through interaction with the central complement convertases and blocking downstream effector functions (Rooijakkers et al. 2007; Rooijakkers and van Strijp 2007). In addition, SaPIov2 encodes two proteins both with 91% amino acid identity to putative regulators encoded by staphylococcal plasmids that confer fusidic acid resistance (O'Neill et al. 2007).

Of the three prophages identified in the genome of strains ED133, φSaov1, highly resembles the φ12 family of staphylococcal phages (Iandolo et al. 2002) and is integrated at ~0.4 Mb in the chromosome at the same site as φSaBov of bovine strain RF122 (Herron-Olson et al. 2007). ΦSaov1 is 45,839 bp in size and includes 69 predicted genes involved in classical phage functions such as integration, capsid and holin formation, and terminase activity (fig. 2*B*). Also encoded by φSaov1 is a putative lipoprotein that shares 95% identity with the putative lipoprotein encoded by φPVL108 (Ma et al. 2006). φSaov2 is 40,345p in length and is integrated at the same chromosomal site in the iron surface determinant (Isd) operon as φNM2 in strain Newman (Bae et al. 2006). φSaov2 encodes an integrase with 94% identity and a lysin with 95% identity to proteins made by φNM2 and encodes several structural proteins related to proteins made by φNM3 (Bae et al. 2006). φSaov2 contains a total of 55 genes, many of which have homology to genes found in several other previously identified staphylococcal phages including φMu50, φPV83, φPVL108, and φSLT, indicating that it is the result of extensive recombination between phages. Of note, φSaov2 encodes four genes that have no homologs among other staphylococcal phages, and a novel superantigen variant (SEA-ov) with 87% amino acid identity to the human-specific

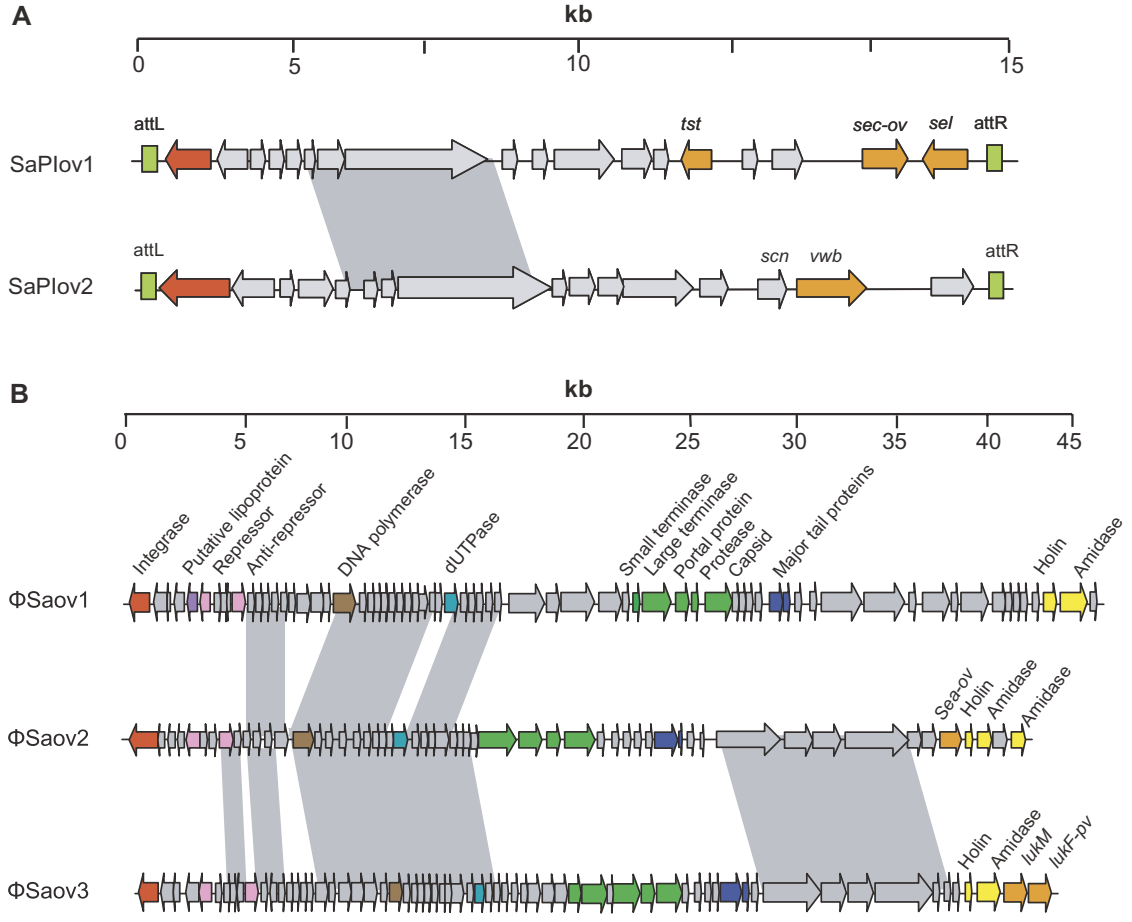

Fig. 2.—Ovine *Staphylococcus aureus* strain ED133 contains novel MGE including staphylococcal pathogenicity islands (SaPIs) (*A*) and prophages (*B*) encoding putative virulence factors. Regions of nucleotide identity (>94%) between MGE are indicated by gray shading, genes in different MGE with the same predicted function have the same color, genes in gray encode hypothetical proteins of unknown function, and green squares denote SaPI attachment sites.

staphylococcal enterotoxin A (SEA) encoded by the β-hemolysin converting phage family (Bae et al. 2006). The third phage identified in the ED133 genome, φSaov3, is 42,947 bp in size and has extensive homology with the ED133 prophage, φSaov2, including the regions specific for the replication machinery, head, and tail proteins (fig. 2*B*) but has a distinct integrase gene and chromosomal insertion site (~2 Mb). φSaov3 also contains regions of homology with the bovine-associated phage, φP83, and human strain-associated phiPVL (Zou et al. 2000) including genes encoding proteins with 99–100% identity to the integrase and to the bicomponent leukotoxin LukM/LukF-PV which is encoded in a remnant phage of 4,465 bp in strain RF122 at a distinct chromosomal site (Herron-Olson et al. 2007). Importantly, previous studies have demonstrated that the LukM/LukF-PV has enhanced activity for bovine leukocytes, consistent with an important role for φSaov3 in ruminant-specific disease pathogenesis (Barrio et al. 2006). Finally, seven copies of insertion element IS1272 (Saov*0030-31*, Saov*0047-48*, Saov*0463-464*, Saov*1809-1810*, Saov*1901-1902*, Saov*2542-2543*, and Saov*2694-2695*) were found, in addition to 6 transposases and 2 truncated transposases. Of note, the IS1272 family of insertion elements have previously been identified in the human clinical *S. aureus* isolate MRSA252, which is proposed to have engaged in horizontal gene transfer events with bovine strains of *S. aureus* (Brody et al. 2008).

Comparative Genomic Hybridization Indicates the Existence of a CC133 Lineage-Specific Accessory Genome

In order to examine the distribution of novel MGE identified in the genome of ED133 among ruminant *S. aureus* isolates and to investigate the variation in gene content among ruminant-associated lineages in general, we carried out comparative genome hybridization analysis of 29 *S. aureus* isolates of bovine, ovine, and caprine origin representative of the genotypic diversity within the major ruminant

lineages CC133, CC97, and CC151, and representative isolates of genotypes less commonly associated with ruminant infections including, ST126, ST130, ST30, and ST39 (supplementary table S1, Supplementary Material online and fig. 3). In addition, the distribution of novel MGE was examined among a panel of 14 human, bovine, and avian *S. aureus* whole-genome sequences (supplementary table S3, Supplementary Material online and fig. 3). The *S. aureus* microarray employed was representative of the genomes of seven strains of human clinical origin (Lindsay et al. 2006) in addition to the bovine strain RF122 (AJ938182) and was updated with sequences unique to ovine strain ED133 identified in the current study. Among the 3,700 *S. aureus*, ORFs represented on the microarray, 1,938 were shared among all strains of the three ruminant lineages CC97 and CC151 and CC133 including 44 genes that were not found among human strains represented on the microarray. Fully annotated microarray data has been deposited in BμG@Sbase (accession number: E-BUGS-94; http://bugs.sgul.ac.uk/E-BUGS-94) and also ArrayExpress (accession number: E-BUGS-94). The major regions of difference between strains were represented by MGEs and core variable regions (fig. 3) consistent with previous comparative genomic analyses of *S. aureus* (Lindsay et al. 2006; Ben Zakour et al. 2008; Sung et al. 2008). CGH analysis supported by PCR confirmation of phage integration sites revealed that all CC133 strains had either a complete ΦSaov1 (n = 9 of 13) or related phage element (n = 4 of 13) and CC151 strains all contained a distinct but related phage element (fig. 3) consistent with the identification of a remnant phage in the genome of the ST151 strain RF122 (Herron-Olson et al. 2007). Similarly, φSaov3 was widely distributed in 13 of the 13 CC133 strains, and a related phage was found in all CC151 strains examined. CGH and PCR analysis indicated that 12 of 13 CC133 isolates contained SaPIov1 (sequencing of the *sec* gene in three representative isolates confirmed the presence of the ovine allelic variant of *sec*, data not shown). Of the CC133 isolates, 11 of 13 contained SaPIov2, and an element distinct from but related to SaPIov2 was detected in one additional CC133 isolate, and 1 of the 4 CC97 isolates examined (fig. 3). In contrast, the prophage φSaov2 was unique to strain ED133 strain but a related phage element was present in human strain Newman (fig. 3).

With the exception of φSaov2, the MGE identified in the sequenced ED133 isolate were widely distributed among CC133 isolates from seven different countries in three continents but were not found among human or other ruminant strains. However, MGE related to but distinct from SaPIov1, SaPIov2, φSaov1, and φSaov3 were found among several ruminant strains of the distinct CC151 and CC97 clonal complexes (fig. 3). In contrast to clones of *S. aureus* colonizing poultry that appear to share a common accessory gene pool, our data suggest limited lateral gene transfer between different *S. aureus* clones sharing the same ruminant niche (fig. 3). Recent studies have demonstrated that the common clonal complexes of *S. aureus* contain unique restriction-modification systems that contribute to the inhibition of genetic exchange between some *S. aureus* lineages (Waldron and Lindsay 2006). Consistent with this hypothesis, we discovered that all CC133 isolates contain the same unique *hsdS1* and *hsdS2 Sau*I genes (data not shown) that may contribute to the conservation of the unique accessory genome of CC133 isolates. The wide distribution of φSaov1, φSaov3, SaPIov1, and SaPIov2 among CC133 isolates from three continents and their absence from all sequenced human isolates examined to date (fig. 3), suggests that they may play an important role in the ruminant host tropism of the CC133 lineage.

SaPIov2 Confers the Ability to Coagulate Ruminant Plasma

In order to test the hypothesis that the unique complement of MGE of CC133 *S. aureus* isolates contributes to its ruminant host tropism, we examined the functional activity of the novel pathogenicity island SaPIov2. SaPIov2 encodes a novel allelic variant of the vWbp encoded in the chromosome of all *S. aureus* strains examined to date, which has been demonstrated to have coagulase activity for human and rabbit plasma (Bjerketorp et al. 2002). Inasmuch as small ruminant and bovine ecovars of *S. aureus* have previously been differentiated from human and other animal strains in their unique ability to coagulate plasma from ruminant sources (Devriese 1984), we hypothesized that SaPIov2 may be responsible for this phenotype. First, selected ruminant (n = 11), human (n = 4), and avian (n = 1) isolates were examined for their ability to stimulate coagulation of rabbit and ruminant plasma (fig. 4). Although all strains could promote coagulation of rabbit plasma, presumably through the coagulase activity of the core genome–encoded coagulase enzyme or vWbp, only isolates which contained SaPIov2 or a related SaPI element which also contained the vWbp^{Sov2} gene could stimulate coagulation of plasma from each of cows, sheep, and goats (coagulation data relating to goat plasma is presented; fig. 4). To examine this correlation further, we deleted SaPIov2 from five different CC133 strains including strain ED133, generating strain ED133 ΔSaPIov2, and replaced SaPIov2 with SaPIN1 from strain N315. Furthermore, we transferred SaPIov2 to the laboratory strain RN4220 resulting in strain RN4220-SaPIov2. Only strains carrying SaPIov2 had the capacity to coagulate plasma from ruminants (fig. 4), indicating that SaPIov2 was responsible for the phenotype. Taken together, these data indicate that a novel staphylococcal pathogenicity island SaPIov2 confers a ruminant host-specific coagulase activity to CC133 *S. aureus* strains, presumably through the activity of the variant vWbp^{Sov2}, which represents a defining phenotype of the traditional biotyping scheme of *S. aureus*. The

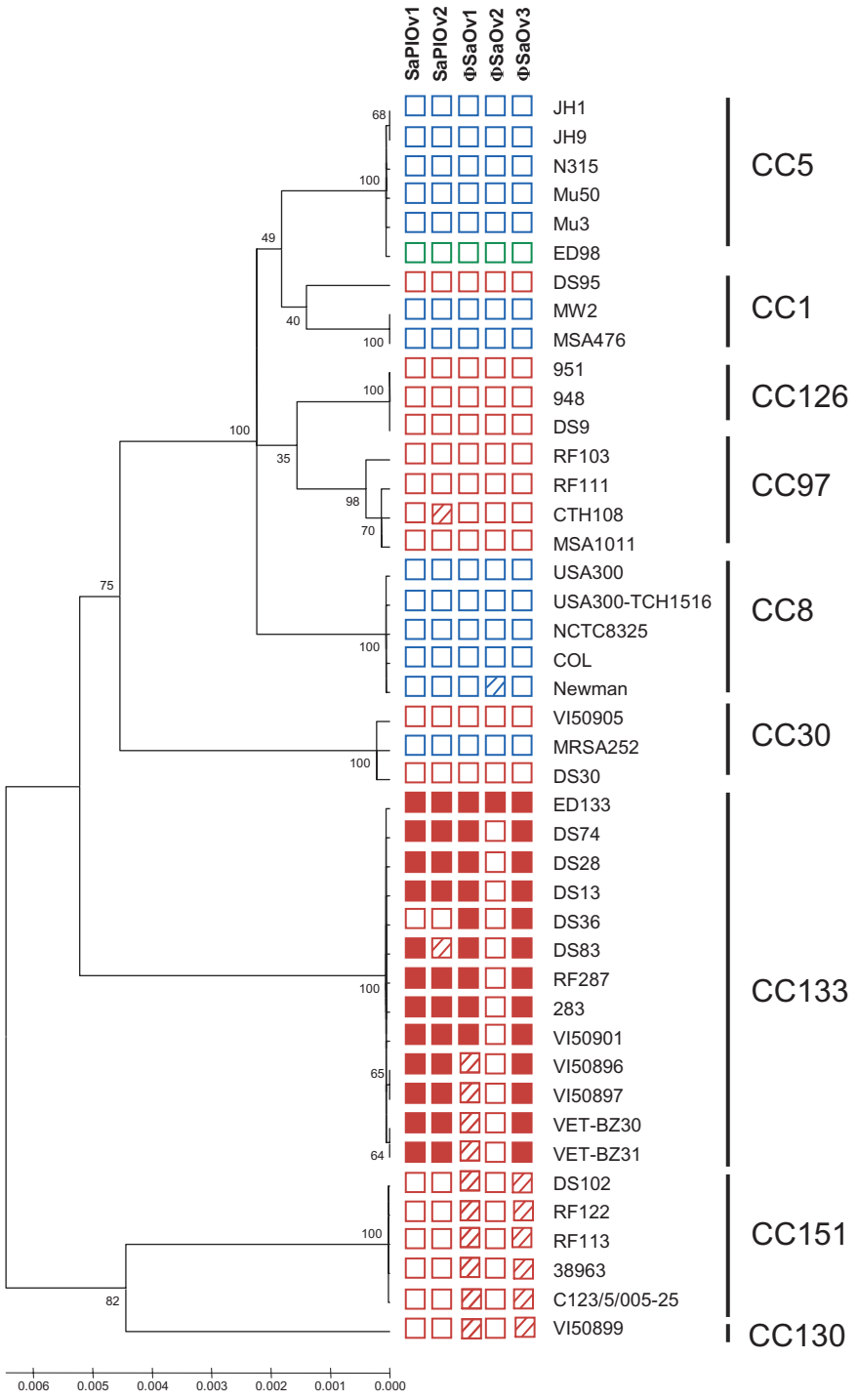

Fig. 3.—Several novel MGE identified in strain ED133 are widely distributed among CC133 isolates but not among human isolates. Strains of ruminant origin associated with the common lineages CC133, CC151, and CC97 and the less common genotypes ST126, ST130, ST30, and ST39, in addition to representative human strains were examined for the presence of MGE identified in strain ED133, determined by a combination of comparative genome hybridization, PCR and/or comparative genome sequence analysis. A filled square indicates the presence of an MGE (confirmed by at least 2 out of 2 positive distinct MGE-specific PCR reactions), an empty space its absence, and a hatched square indicates the presence of a distinct but genetically related element (determined by at least 1 out of 2 positive distinct MGE-specific PCR reaction). Red, blue, and green boxes denote strains of ruminant, human, or avian origin, respectively. A Neighbor-Joining consensus tree inferred from 500 bootstrap replicates was constructed based on the concatenated sequences of STs to indicate the genetic relatedness of the strains examined.

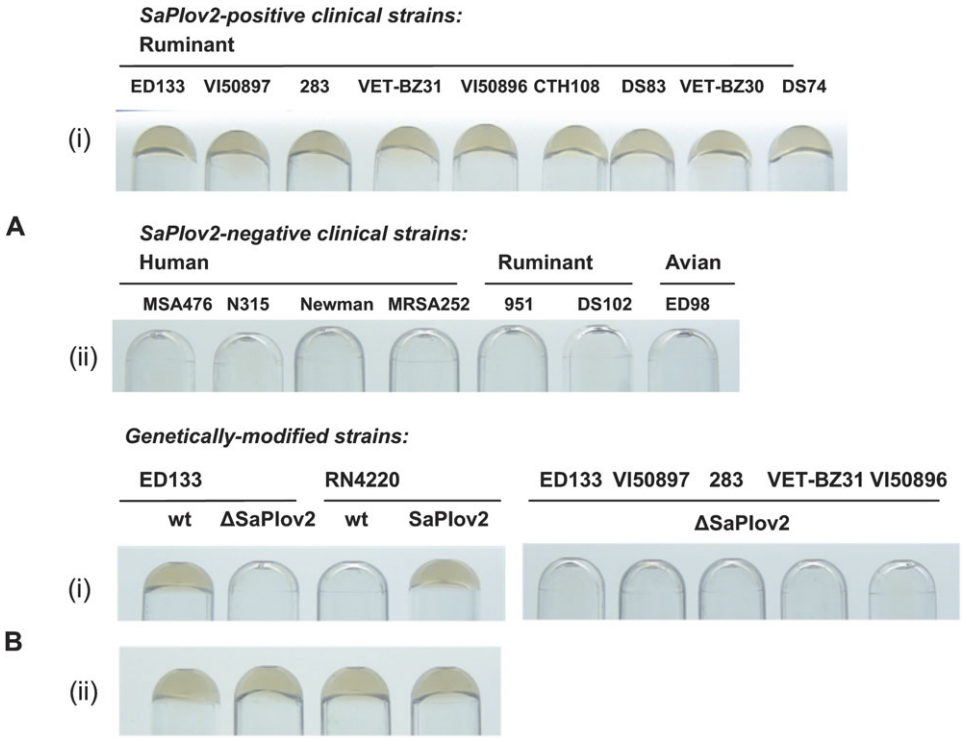

Fig. 4.—SaPIov2 confers the ability to coagulate ruminant plasma. Ruminant plasma coagulase assays represented by inverted test tubes with or without plasma clots indicating positive or negative results, respectively. (*A*) Wild-type clinical isolates of ruminant, human, or avian origin with (i) or without (ii) the novel MGE SaPIov2. (*B*) (i) Ruminant plasma coagulase activity of genetically modified isolates including clinical isolates with SaPIov2 replaced by SaPIN1 from human strain N315, and lab strain RN4220 which has acquired SaPIov2. (ii) Rabbit plasma coagulase activity of selected strains (all strains examined were positive for rabbit plasma coagulase activity, data not shown).

identification of a novel pathogenicity island specific for ruminant strains, with a host-specific virulence-associated function, suggests an important role in the host tropism of the CC133 clonal lineage.

Conclusions

The identification over 60 years ago of unique combinations of phenotypes associated with strains from different host species led to the development of a biotyping scheme for *S. aureus* and represented the first evidence for the existence of host-specialized ecovars of *S. aureus*. Here, we provide evidence that the small ruminant ecovar originated as a result of a host jump from humans and underwent host adaptation through a combination of gene diversification, decay, and horizontal acquisition of MGE encoding virulence proteins with attenuated or enhanced activity in ruminants. The identification of a widely distributed novel pathogenicity island, which confers the ability to coagulate ruminant plasma reveals the basis for a defining phenotype of the classical *S. aureus* biotyping scheme. Our results highlight the central role that MGE play in the adaptation of bacteria to different host species and reveal broad new insights into the molecular basis of *S. aureus* host specialization.

Supplementary Material

Supplementary figure S1 and tables S1–S5 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Acknowledgments

This work was funded by grant BB/D521222/1 from the Biotechnology and Biological Sciences Research Council (to J.R.F.). The Bacterial Microarray Group at St Georges is funded by The Wellcome Trust. We are grateful to Y. Le Loir for providing ED133 (formerly 1174), H. de Lencastre, H. Joergensen, and L. Green for providing ruminant isolates and the J Craig Venter Institute for assistance with genome annotation.

Literature Cited

Aires-de-Sousa M, et al. 2007. Characterization of Staphylococcus aureus isolates from buffalo, bovine, ovine, and caprine milk samples collected in Rio de Janeiro State, Brazil. Appl Environ Microbiol. 73:3845–3849.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics. 164:1229–1236.

Aziz R, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9:75.

Baba T, et al. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet. 359:1819–1827.
Bae T, Baba T, Hiramatsu K, Schneewind O. 2006. Prophages of Staphylococcus aureus Newman and their contribution to virulence. Mol Microbiol. 62:1035–1047.
Barrio MB, Rainard P, Prevost G. 2006. LukM/LukF'-PV is the most active Staphylococcus aureus leukotoxin on bovine neutrophils. Microbes Infect. 8:2068–2074.
Battisti A, et al. 2010. Heterogeneity among methicillin-resistant Staphylococcus aureus from Italian pig finishing holdings. Vet Microbiol. 142:361–366.
Ben Zakour NL, et al. 2008. Genome-wide analysis of ruminant Staphylococcus aureus reveals diversification of the core genome. J Bacteriol. 190:6302–6317.
Bjerketorp J, et al. 2002. A novel von Willebrand factor binding protein expressed by Staphylococcus aureus. Microbiology. 148:2037–2044.
Bjerketorp J, Rosander A, Nilsson M, Jacobsson K, Frykberg L. 2004. Sorting a Staphylococcus aureus phage display library against ex vivo biomaterial. J Med Microbiol. 53:945–951.
Brody T, et al. 2008. Horizontal gene transfers link a human MRSA pathogen to contagious bovine mastitis bacteria. PLoS One. 3:e3074.
Bruckner R. 1997. Gene replacement in Staphylococcus carnosus and Staphylococcus xylosus. FEMS Microbiol Lett. 151:1–8.
Bubeck Wardenburg J, Williams WA, Missiakas D. 2006. Host defenses against Staphylococcus aureus infection require recognition of bacterial lipoproteins. Proc Natl Acad Sci U S A. 103:13831–13836.
Carver TJ, et al. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics. 24:2672–2676.
Carver TJ, et al. 2005. ACT: the Artemis Comparison Tool. Bioinformatics. 21:3422–3423.
Cheng AG, et al. 2009. Genetic requirements for Staphylococcus aureus abscess formation and persistence in host tissues. FASEB J. 23:3393–3404.
Clarke SR, Foster SJ. 2006. Surface adhesins of Staphylococcus aureus. Adv Microb Physiol. 51:187–224.
Cooper JE, Feil EJ. 2006. The phylogeny of Staphylococcus aureus—which genes make the best intra-species markers? Microbiology. 152:1297–1305.
Corrigan RM, Miajlovic H, Foster TJ. 2009. Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. BMC Microbiol. 9:22.
Deringer JR, Ely RJ, Monday SR, Stauffacher CV, Bohach GA. 1997. Vbeta-dependent stimulation of bovine and human T cells by host-specific staphylococcal enterotoxins. Infect Immun. 65:4048–4054.
Devriese LA. 1984. A simplified system for biotyping Staphylococcus aureus strains isolated from animal species. J Appl Bacteriol. 56:215–220.
Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.
Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.
Eppinger M, et al. 2006. Who ate whom? Adaptive Helicobacter genomic changes that accompanied a host jump from early humans to large felines. PLoS Genet. 2:e120.
Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.
Feil EJ, et al. 2003. How clonal is Staphylococcus aureus? J Bacteriol. 185:3307–3316.
Fitzgerald JR, Meaney WJ, Hartigan PJ, Smyth CJ, Kapur V. 1997. Fine-structure molecular epidemiological analysis of Staphylococcus aureus recovered from cows. Epidemiol Infect. 119: 261–269.
Fitzgerald JR, et al. 2001. Characterization of a putative pathogenicity island from bovine Staphylococcus aureus encoding multiple superantigens. J Bacteriol. 183:63–70.
Gordon D. 2003. Viewing and editing assembled sequences using Consed. Curr Protoc Bioinformatics. Chapter 11:Unit11.2
Guinane CM, et al. 2008. Pathogenomic analysis of the common bovine Staphylococcus aureus clone (ET3): emergence of a virulent subtype with potential risk to public health. J Infect Dis. 197: 205–213.
Harris SR, et al. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 327:469–474.
Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22:160–174.
Herron-Olson L, Fitzgerald JR, Musser JM, Kapur V. 2007. Molecular correlates of host specialization in Staphylococcus aureus. PLoS One. 2:e1120.
Herron LL, et al. 2002. Genome sequence survey identifies unique sequences and key virulence genes with unusual rates of amino acid substitution in bovine Staphylococcus aureus. Infect Immun. 70:3978–3981.
Iandolo JJ, et al. 2002. Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of Staphylococcus aureus 8325. Gene. 289:109–118.
Ito T, et al. 2001. Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother. 45:1323–1336.
Jorgensen HJ, Mork T, Hogasen H, Rorvik LM. 2005. Enterotoxigenic Staphylococcus aureus in bulk milk in Norway. J Appl Microbiol. 99:158–166.
Kapur V, Sischo WM, Greer RS, Whittam TS, Musser JM. 1995. Molecular population genetic analysis of Staphylococcus aureus recovered from cows. J Clin Microbiol. 33:376–380.
Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimates of the evolutionary tree topologies from sequence data, and the branching order in Hominoidea. J Mol Evol. 29: 170–179.
Kosakovsky Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.
Lee PK, et al. 1992. Nucleotide sequences and biologic properties of toxic shock syndrome toxin 1 from ovine- and bovine-associated Staphylococcus aureus. J Infect Dis. 165:1056–1063.
Lefebure T, Stanhope MJ. 2007. Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. Genome Biol. 8:R71.
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.
Lowder BV, et al. 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. Proc Natl Acad Sci U S A. 106:19545–19550.
Ma XX, Ito T, Chongtrakool P, Hiramatsu K. 2006. Predominance of clones carrying Panton-Valentine leukocidin genes among methicillin-resistant Staphylococcus aureus strains isolated in Japanese hospitals from 1979 to 1985. J Clin Microbiol. 44:4515–4527.

Madison RR. 1935. Fibrinolytic staphylococci. Proc Sac Exp Biol. 33:309.
Minnett FC. 1936. Staphylococci from animals with particular reference to toxin production. J Pathol Bacteriol. 42:247–259.
Moreillon P, et al. 1995. Role of Staphylococcus aureus coagulase and clumping factor in pathogenesis of experimental endocarditis. Infect Immun. 63:4738–4743.
Mwangi MM, et al. 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Proc Natl Acad Sci U S A. 104:9451–9456.
Ochman H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. Mol Biol Evol. 20:2091–2096.
O'Neill AJ, Larsen AR, Skov R, Henriksen AS, Chopra I. 2007. Characterization of the epidemic European fusidic acid-resistant impetigo clone of Staphylococcus aureus. J Clin Microbiol. 45:1505–1510.
Overbeek R, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33:5691–5702.
Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007. Genes under positive selection in Escherichia coli. Genome Res. 17:1336–1343.
Raulf M, Alouf JE, Konig W. 1990. Effect of staphylococcal delta-toxin and bee venom peptide melittin on leukotriene induction and metabolism of human polymorphonuclear granulocytes. Infect Immun. 58:2678–2682.
Robinson DA, Monk AB, Cooper JE, Feil EJ, Enright MC. 2005. Evolutionary genetics of the accessory gene regulator (agr) locus in Staphylococcus aureus. J Bacteriol. 187:8312–8321.
Rodgers JD, McCullagh JJ, McNamee PT, Smyth JA, Ball HJ. 1999. Comparison of Staphylococcus aureus recovered from personnel in a poultry hatchery and in broiler parent farms with those isolated from skeletal disease in broilers. Vet Microbiol. 69:189–198.
Rooijakkers SH, et al. 2007. Staphylococcal complement inhibitor: structure and active sites. J Immunol. 179:2989–2998.
Rooijakkers SH, et al. 2006. Early expression of SCIN and CHIPS drives instant immune evasion by Staphylococcus aureus. Cell Microbiol. 8:1282–1293.
Rooijakkers SH, van Strijp JA. 2007. Bacterial complement evasion. Mol Immunol. 44:23–32.
Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16:944–945.
Schmitz FJ, Veldkamp KE, Van Kessel KP, Verhoef J, Van Strijp JA. 1997. Delta-toxin from Staphylococcus aureus as a costimulator of human neutrophil oxidative burst. J Infect Dis. 176:1531–1537.
Schoen C, Joseph B, Claus H, Vogel U, Frosch M. 2007. Living in a changing environment: insights into host adaptation in Neisseria meningitidis from comparative genomics. Int J Med Microbiol. 297:601–613.
Shriner D, Nickle DC, Jensen MA, Mullins JI. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet Res. 81:115–121.
Sloane R, et al. 1991. A toxic shock syndrome toxin mutant of Staphylococcus aureus isolated by allelic replacement lacks virulence in a rabbit uterine model. FEMS Microbiol Lett. 62:239–244.
Smith EM, et al. 2005. Multilocus sequence typing of intercontinental bovine Staphylococcus aureus isolates. J Clin Microbiol. 43: 4737–4743.
Smyth DS, et al. 2009. Molecular genetic typing reveals further insights into the diversity of animal-associated Staphylococcus aureus. J Med Microbiol. 58:1343–1353.
Smyth DS, et al. 2010. Population structure of a hybrid clonal group of methicillin-resistant Staphylococcus aureus, ST239-MRSA-III. PLoS One. 5:e8582.
Stinear TP, et al. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. Genome Res. 17:192–200.
Strimmer K. 2008. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics. 24: 1461–1462.
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.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.
Tcherepanov V, Ehlers A, Upton C. 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics 7:150.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Ubeda C, et al. 2005. Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. Mol Microbiol. 56:836–844.
van Hijum SA, Zomer AL, Kuipers OP, Kok J. 2005. Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. Nucleic Acids Res. 33:W560–W566.
van Leeuwen WB, et al. 2003. Multilocus sequence typing of Staphylococcus aureus with DNA array technology. J Clin Microbiol. 41:3323–3326.
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.
Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. Curr Protoc Bioinformatics. Chapter 6:Unit 6.4
Wren BW. 2000. Microbial genome analysis: insights into virulence, host adaptation and evolution. Nat Rev Genet. 1:30–39.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17:32–43.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22: 1107–1118.
Zou D, Kaneko J, Narita S, Kamio Y. 2000. Prophage, phiPV83-pro, carrying panton-valentine leukocidin genes, on the Staphylococcus aureus P83 chromosome: comparative analysis of the genome structures of phiPV83-pro, phiPVL, phi11, and other phages. Biosci Biotechnol Biochem. 64:2631–2643.

Associate editor: Hidemi Watanabe