Stable antibiotic resistance and rapid human adaptation in livestock-associated MRSA

Marta Matuszewska[1,*], Gemma G. R. Murray[1,*,†], Xiaoliang Ba[1], Rhiannon Wood[1], Mark A. Holmes[1] and Lucy A. Weinert[1]

1. Department of Veterinary Medicine, University of Cambridge, Cambridge, CB3 0ES, UK.

* These authors contributed equally and should be acknowledged as joint first authors.

† Corresponding author. Email: ggrmurray@gmail.com.

Keywords: *Staphylococcus aureus*, MRSA, livestock-associated MRSA clonal-complex 398, host-switching, zoonosis, antibiotic resistance, mobile genetic elements, Tn*916* transposon, SCC*mec*, ϕSa3 prophage

Abstract

Mobile genetic elements (MGEs) are agents of horizontal gene transfer in bacteria, but can also be vertically inherited by daughter cells. Establishing the dynamics that led to contemporary patterns of MGEs in bacterial genomes is central to predicting the emergence and evolution of novel and resistant pathogens. Methicillin-resistant *Staphylococcus aureus* (MRSA) clonal-complex (CC) 398 is the dominant MRSA in European livestock and a growing cause of human infections. Previous studies have identified three categories of MGEs whose presence or absence distinguishes livestock-associated CC398 from a closely related and less antibiotic-resistant human-associated population. Here we fully characterise the evolutionary dynamics of these MGEs using a collection of 1,180 CC398 genomes, sampled from livestock and humans, over 27 years. We find that the emergence of livestock-associated CC398 coincided with the acquisition of a Tn*916* transposon carrying a tetracycline resistance gene, which has been stably inherited for 57 years. This was followed by the acquisition of a type V SCC*mec* that carries methicillin, tetracycline and heavy metal resistance genes, which has been maintained for 35 years, with occasional truncations and replacements with type IV SCC*mec*. In contrast, a class of prophages that carry a human immune evasion gene cluster and that are largely absent from livestock-associated CC398, have been repeatedly gained and lost in both human- and livestock-associated CC398. These contrasting dynamics mean that when livestock-associated MRSA is transmitted to humans, adaptation to the human host outpaces loss of antibiotic resistance. In addition, the stable inheritance of resistance-associated MGEs suggests that the impact of ongoing reductions in antibiotic and zinc oxide use in European farms on livestock-associated MRSA will be slow to be realised.

Introduction

Mobile genetic elements (MGEs) play an important role in the evolution of bacterial pathogens. They can move rapidly between bacterial genomes, but can also be vertically inherited through stable integration into a host genome. As MGEs often carry genes associated with virulence and antibiotic resistance (Frost et al. 2005; Rankin et al. 2011), an understanding of the drivers and barriers to their acquisition and maintenance is central to predicting the emergence and evolution of novel and resistant pathogens (Brockhurst et al. 2019).

The emergence and evolution of methicillin-resistant *Staphylococcus aureus (*MRSA) across different ecological niches and host species is associated with the horizontal transfer of MGEs. Methicillin resistance is carried by the staphylococcal cassette chromosome element SCC*mec*, and additional MGEs carry resistance to other antibiotics, virulence factors and host-specific adaptations (Hanssen and Ericson Sollid 2006; Jamrozy et al. 2017; Haag et al. 2019; Turner et al. 2019; Matuszewska et al. 2020). While most MRSA clonal complexes (CCs) show an association with specific MGEs, their dynamics are not widely understood, leading to a gap in our understanding of the adaptive potential of *S. aureus* CCs.

Intensification combined with high levels of antibiotic use in farming has led to particular concerns about livestock as reservoirs of antibiotic-resistant human infections(WHO 2015). CC398 has become the dominant MRSA in European livestock. Its rise has been particularly evident in Danish pig farms where the proportion of MRSA-positive herds has increased from <5% in 2008 to 90% in 2018 (Sieber et al. 2018; DANMAP 2019), but it has also been observed in other European countries, and other livestock species (Lekkerkerk et al. 2015; Islam et al. 2017; Anjum et al. 2019). Livestock-associated (LA) MRSA CC398 has been associated with increasing numbers of human infections, in both people with and without direct contact with livestock (Larsen et al. 2017; van Alen et al. 2017; Sieber et al. 2019). Understanding the emergence and success of CC398 in European livestock and its capacity to infect the human host is integral to managing the risk that it, and other livestock-associated pathogens, pose to public health.

Previous studies have used genome sequences to reconstruct the evolutionary history of CC398 (Price et al. 2012; Ward et al. 2014; Silva et al. 2017). They identified a largely methicillin-resistant and livestock-associated clade of CC398, that falls as either sister to (Ward et al. 2014) or within (Price et al. 2012; Silva et al. 2017) a largely methicillin-sensitive

and human-associated clade. Through comparing the genomes of isolates from livestock-associated and human-associated CC398, these studies concluded that the emergence of CC398 in livestock was associated with both the acquisition of antibiotic resistance genes and the loss of genes associated with human immune evasion. While these genes are known to be carried on three categories of MGEs (Tn*916* conjugative transposons, SCC*mec* and ϕSa3 prophages), little is known about the dynamics of these MGEs within CC398. Here, we undertake a comprehensive reconstruction of the evolutionary dynamics of these MGEs. We find that while their patterns of presence/absence all show a strong association with the transition to livestock, this is the result of contrasting dynamics. These dynamics can inform predictions about the risk posed by LA-MRSA spillover infections in humans, and the resilience of antibiotic resistance in LA-MRSA to ongoing changes in antibiotic use in farming.

Results

Livestock-associated CC398 emerged between 1957 and 1970

We collected and assembled publicly available whole-genome sequence data from CC398, and sequenced five isolates recently sampled from pig farms in the UK. Our collection includes high-quality whole genome assemblies of 1,180 isolates (including 43 complete reference genomes). This collection spans 15 host species (including humans, pigs, cows, chickens, turkeys and horses), 28 countries (across Europe, America, Asia, and Australasia), and 27 years (1992 to 2018) (Figure 1-figure supplement 1, Supplementary File 1).

We constructed a recombination-stripped maximum likelihood phylogeny of CC398 using reference-mapped assemblies of our collection. We rooted the phylogeny with outgroups from four other *S. aureus* sequence types (STs) in four separate reconstructions. We also constructed a phylogeny from a recombination-stripped concatenated alignment of core genes extracted from *de novo* assemblies, with a midpoint rooting. These reconstructions consistently returned the same topology, and one that is described in previous studies: a livestock-associated clade of CC398 (704 isolates) that falls within a more diverse, and other largely human-associated clade (476 isolates) (Figure 1a and Figure 1- figure supplement 2) (Price et al. 2012; Silva et al. 2017).

We used the temporal structure in our collection to date the origin of the livestock-associated clade. Due to the size of our collection, we constructed dated phylogenies from three

subsampled data sets, each of which includes 250 isolates. Our estimates of the evolutionary rate were consistent across all reconstructions (1.1-1.6 x10^{-6} subs/site/year), and similar to estimates from previous studies of CC398 (Ward et al. 2014) and other *S. aureus* CCs (Hsu et al. 2015). This led to an estimate of the origin of the livestock-associated clade of approximately 1964 (95% CI: 1957-1970) (Figure 1b, Figure 1- figure supplement 4 and Figure 1- figure supplement 4).

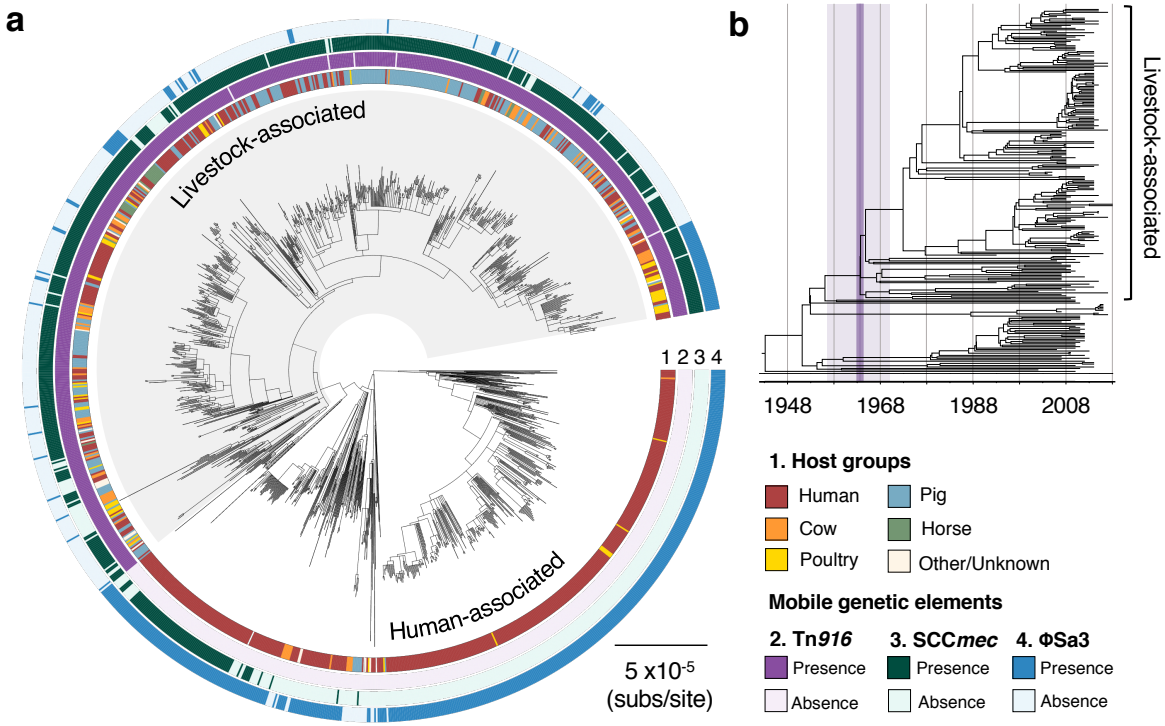


Figure 1. The transition to livestock-association in the 1960s was accompanied by changes in the frequencies of three mobile genetic elements. (a) A maximum likelihood phylogeny of 1,180 isolates of CC398, rooted using an outgroup from ST291. Grey shading indicates the livestock-associated clade. Outer rings describe (1) the host groups isolates were sampled from, and the presence of three MGEs: (2) a Tn*916* transposon carrying *tetM*, (3) a SCC*mec* carrying *mecA*, and (4) a ϕSa3 prophage carrying a human immune evasion gene cluster. (b) A dated phylogeny of a sample of 250 CC398 isolates, that shows livestock-associated CC398 originated around 1964 (95% HPD: 1957-1970).

The transition to livestock-association is associated with changes in the frequencies of three mobile genetic elements with very different dynamics

Comparisons of the genomes of isolates from human- and livestock-associated CC398, have previously indicated that the transition to livestock was associated with the acquisition

of genes associated with both tetracycline and methicillin resistance (*tetM* and *mecA*), and the loss of genes associated with human immune evasion (the immune evasion gene cluster) (Price et al. 2012). Our analyses of this larger collection are broadly consistent with this. We find that the genes whose presence most strongly distinguishes isolates from the human- and livestock-associated groups are associated with three categories of MGEs: (1) a Tn*916* transposon carrying *tetM*; (2) SCC*mec* carrying *mecA*; and (3) ϕSa3 prophages carrying a human immune evasion gene cluster (Figure 1a, Figure 1- figure supplement 5, Supplementary File 2). Genes associated with the Tn*916* transposon and SCC*mec* elements are more common in the livestock-associated clade, while the reverse is true of genes associated with ϕSa3 prophages.

1. *Stable maintenance of a Tn*916 *transposon carrying* tetM

We identified a contiguous assembly of a Tn*916* transposon carrying *tetM* in 699/704 isolates in our collection of livestock-associated CC398 (Figure 2a, Supplementary File 3 and Supplementary File 4) (de Vries et al. 2009; Roberts and Mullany 2009). Several lines of evidence indicate that the presence of the Tn*916* transposon in livestock-associated CC398 is the result of a single acquisition event, followed by stable inheritance. First, the location of the transposon in the genome of livestock-associated CC398 is conserved (Tn*916* is always found next to the same core gene; WP_000902814 in the published annotation of S0385). Second, an alignment of the coding regions of this element (extracted from *de novo* assemblies) shows a similar average nucleotide diversity to core genes (Figure 2d). Third, a phylogeny constructed from the genes in this element is entirely congruent with the phylogeny of livestock-associated CC398 (Figure 2b,c).

Our analyses indicate that the Tn*916* transposon has been maintained in livestock-associated CC398 since its origin, and therefore for around 57 years (Figure 1b). Nevertheless, its absence from 5/704 livestock-associated CC398 isolates in our collection suggests that it remains capable of excision (this is consistent with experimental studies that have shown that Tn*916* in CC398 is a functional conjugative transposon; de Vries et al. 2009). None of the genes associated with this element are present in these five isolates, and in two we were able to identify an intact integration site in the assembled genome (Figure 2-figure supplement 1). These five isolates are broadly distributed across the livestock-associated clade, and not linked to a particular host species or geographic location.

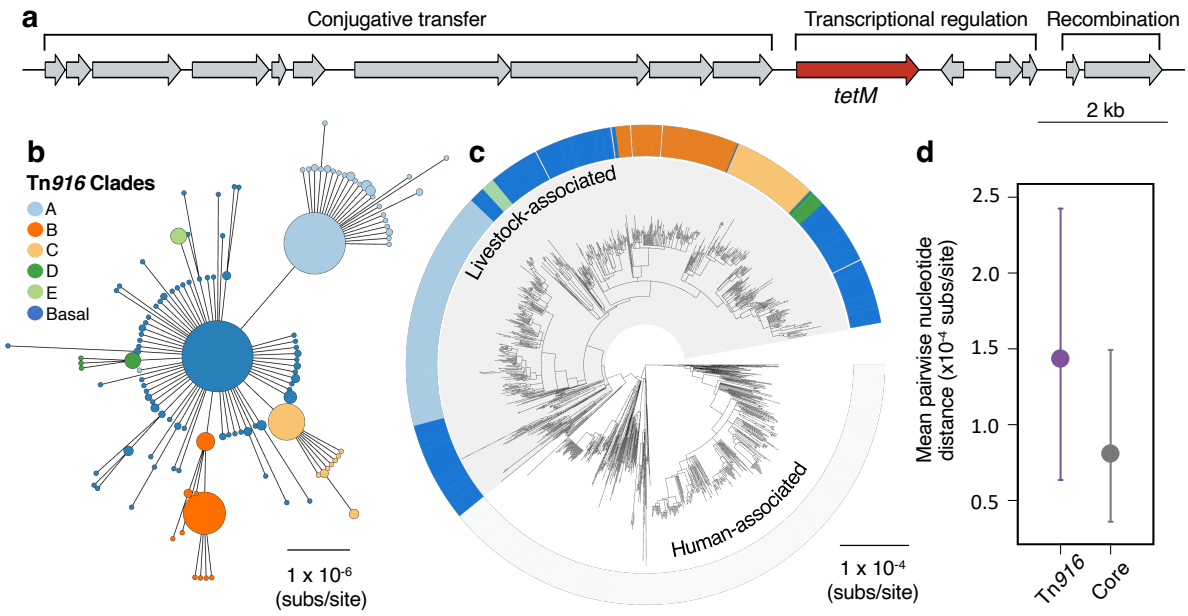


Figure 2. A Tn*916* transposon carrying *tetM* has been stably maintained by livestock-associated CC398 since its origin. (a) A gene map of the Tn*916* transposon in CC398 **(**based on reference genome 1_1439), with annotations based on previous studies(de Vries et al. 2009; Roberts and Mullany 2009). (b) A minimum spanning tree of the element based on a concatenated alignment of all genes shown in (a). Points represent groups of identical elements, with point size correlated with number of elements on a log-scale, and colours representing well-supported clades (>70 bootstrap support in a maximum likelihood phylogeny) that include >10 elements (smaller clades are incorporated into their basal clade). (c) These clades are annotated onto the CC398 phylogeny as an external ring. (d) Mean pairwise nucleotide distance between isolates carrying the Tn*916* transposon based on genes in the Tn*916* transposon and core genes, using bootstrapping to estimate error (see methods for details).

*2. More variable maintenance of a SCC*mec *carrying* mecA*,* tetK *and* czrC

Previous studies suggest that LA-MRSA CC398 emerged from human-associated MSSA (Price et al. 2012). However, the presence of SCC*mec* elements in recently sampled human-associated CC398 isolates that fall basal to the livestock-associated clade, including clinical isolates from China (He et al. 2018; Zou et al. 2022), Denmark (Møller et al. 2019), and New Zealand (Silva et al. 2017), makes the association between methicillin-resistance and livestock-association less clear (Figure 1a).

SCC*mec* elements in *S. aureus* are categorised into several types (Hanssen and Ericson Sollid 2006). Consistent with previous studies of CC398 (e.g. Price et al. 2012), we observe

both type V (76%) and type IV (21%) in CC398 (we were unable to confidently type the remaining 3%; Figure 3c, Figure 3-figure supplement 1, Supplementary File 3 and Supplementary File 5). Most of the type V SCC*mec* elements belong to the subtype Vc previously described in livestock-associated CC398 (Li et al. 2011; Price et al. 2012; Vandendriessche et al. 2014). This element includes two additional resistance genes: *tetK* (tetracycline resistance) and *czrC* (heavy metal resistance) (Figure 3a, Supplementary File 6). We identified a full-length version of this element in 335 genomes (including some of the most recent isolates in our collection, sampled from UK pig farms in 2018) and shorter type V elements in 204 genomes. Full-length versions are only observed in the livestock-associated clade, while shorter versions are found in both livestock- and human-associated groups of CC398. Shorter versions often lacked *tetK* (n=90) and *czrC* (n=117) genes (Figure 3-figure supplement 2). Type IV elements are only observed in the livestock-associated clade. They include subtypes IVa and IVc, and all only carry a single resistance gene: *mecA*.

While type Vc is the most common SCC*mec* in our collection of livestock-associated CC398, this largely reflects isolates from pigs. Type Vc SCC*mec* is much more common in pigs (77% of isolates; n=286) than type IV SCC*mec* (7% of isolates). In cows (n=74) the difference is reduced: 45% are type Vc and 32% are type IV. And in isolates from other animal species (n=94) type IV elements (55%) are more common than type Vc (33%).

Diversity within the type Vc SCC*mec* element indicates that a full-length type Vc SCC*mec* was acquired once by livestock-associated CC398, and has been maintained within CC398 largely through vertical transmission. First, low nucleotide diversity within full-length versions of the element is consistent with 329/335 sharing a recent origin common within livestock-associated CC398 (Figure 3d). Second, patterns of diversity are largely congruent with the core genome phylogeny, consistent with vertical inheritance (Figure 3b,c). Third, low nucleotide diversity within genes shared across full- and shorter-length versions of the element are consistent with most shorter-length versions being the result of deletion within livestock-associated CC398 (Figure 3-figure supplement 2). In contrast, diversity in the type IV elements in livestock-associated CC398 supports four independent acquisitions from outside of CC398 (Figure 3-figure supplement 3). Similar to the type Vc element, once acquired, these elements tend to be maintained.

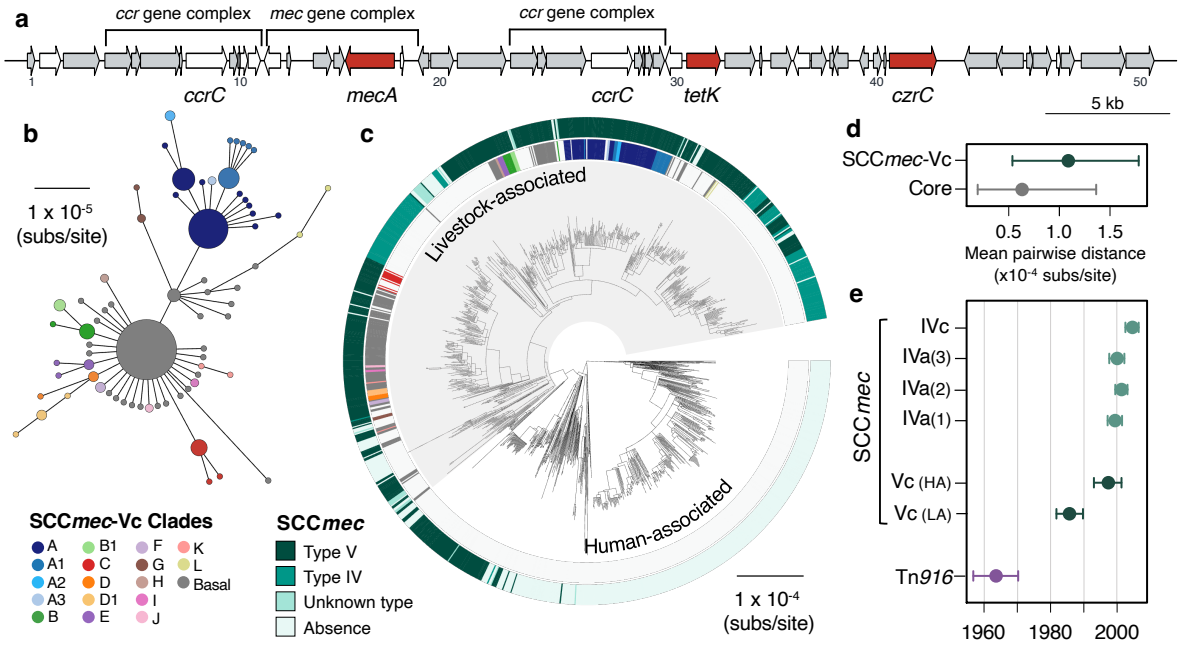


Figure 3. A type V SCC*mec* has been maintained since the 1980s, with occasional replacements. (a) A gene map of the type Vc *SCCmec* element in CC398 (using the 1_1439 reference strain), with annotations from previous studies (Li et al. 2011; Vandendriessche et al. 2014). Genes in white were excluded from analyses of diversity within the element due to difficulties in distinguishing homologues. (b) A minimum-spanning tree of the type Vc *SCCmec* element based on a concatenated alignment of the genes (grey and red) in (a). Points represent groups of identical elements, point size correlates with group size on a log-scale, and colours represent well-supported clades (>70 bootstrap support in a maximum likelihood phylogeny). (c) Well-supported clades and SCC*mec* type are annotated on the CC398 phylogeny in external rings. (d) Mean pairwise nucleotide distance between isolates carrying the SCC*mec* type Vc based on genes in the SCC*mec* type Vc and core genes, with error estimated by bootstrapping (see methods for details). (e) Acquisition dates for different SCC*mec* elements and Tn*916* inferred from an ancestral state reconstruction over the dated phylogeny in 1 (b). Dates for type Vc are shown for both livestock and human-associated CC398.

While the SCC*mec* elements carried by human-associated CC398 are always type V, 70/80 fall within a single clade from a hospital outbreak in Denmark in 2016 (Møller et al. 2019). They show a truncation relative to the full-length type Vc that is also observed in 3 isolates from livestock-associated CC398 (leading to the absence of *czrC*, Figure 3-figure supplement 2). Pairwise nucleotide distances between these 70 human-associated CC398 elements and the full-length type Vc in livestock-associated CC398 are consistent with a recent common ancestor within CC398. In contrast, nucleotide diversity within the other 10 type V SCC*mec* in human-associated CC398 and distances from the livestock-associated

CC398 type Vc indicate multiple independent acquisitions from outside of CC398 (Figure 3-figure supplement 2).

Using our dated phylogeny of CC398 and categorisation of SCC*mec* based on diversity within the elements, we inferred the dynamics of gain and loss within CC398 (Figure 3e and Figure 3-figure supplement 4). These reconstructions consistently estimated that the type Vc SCC*mec* had been acquired by livestock-associated CC398 by around 1986 (95% CI: 1982-1990), and has therefore been maintained within livestock-associated CC398 for around 35 years. While the diversity within this element indicates a single acquisition by CC398, these reconstructions indicate multiple gains, likely reflecting horizontal transmission within CC398. They also indicate that the acquisition by livestock-associated CC398 predated the acquisition by human-associated CC398, consistent with transmission from livestock-associated CC398 to human-associated CC398. In contrast, type IV elements show evidence of several more recent acquisitions in relatively quick succession between 1997 and 2004 (Figure 3e and Figure 3-figure supplement 4).

Together, our analyses are consistent with LA-MRSA emerging from human-associated MSSA and acquiring the type Vc SCC*mec* element following its initial diversification. The element has been stably maintained (particularly in pigs), although with several exceptions - including deletions of parts of the element and replacements with smaller type IV SCC*mec*.

3. Loss of a ϕSa3 prophage carrying a human immune evasion gene cluster, but with frequent reacquisition

In contrast to the maintenance of Tn*916* and SCC*mec* type V in the livestock-associated clade, the ϕSa3 prophage is highly dynamic. ϕSa3 prophages carry human immune evasion gene clusters that include a variable set of functional genes that encode human-specific virulence factors, including *sak, chp*, *scn*, *sea* and *sep* (Gladysheva et al. 2003; Postma et al. 2004; Rooijakkers et al. 2005; van Wamel et al. 2006; Thammavongsa et al. 2015; van Alen et al. 2018). They are temperate prophages that primarily integrate into the *hlb* gene of *S. aureus*. While ϕSa3 prophages are present in 88% of the human-associated CC398 isolates in our collection, we find that this is not a consequence of a stable association between CC398 and one prophage. Nucleotide diversity within genes shared across ϕSa3 prophages (Figure 4, Supplementary File 3), suggest at least seven (but likely more) acquisitions of an ϕSa3 prophage into human-associated CC398 from outside of CC398. The set of functional genes carried by these elements indicate that the elements within

human-associated CC398 include types C (n=285), B (n=111), E (n=35), A (n=4) and D (n=4), and those carried by livestock-associated CC398 isolates include types B (n=84), E (n=8) and A (n=5) (van Wamel et al. 2006; van Alen et al. 2018).

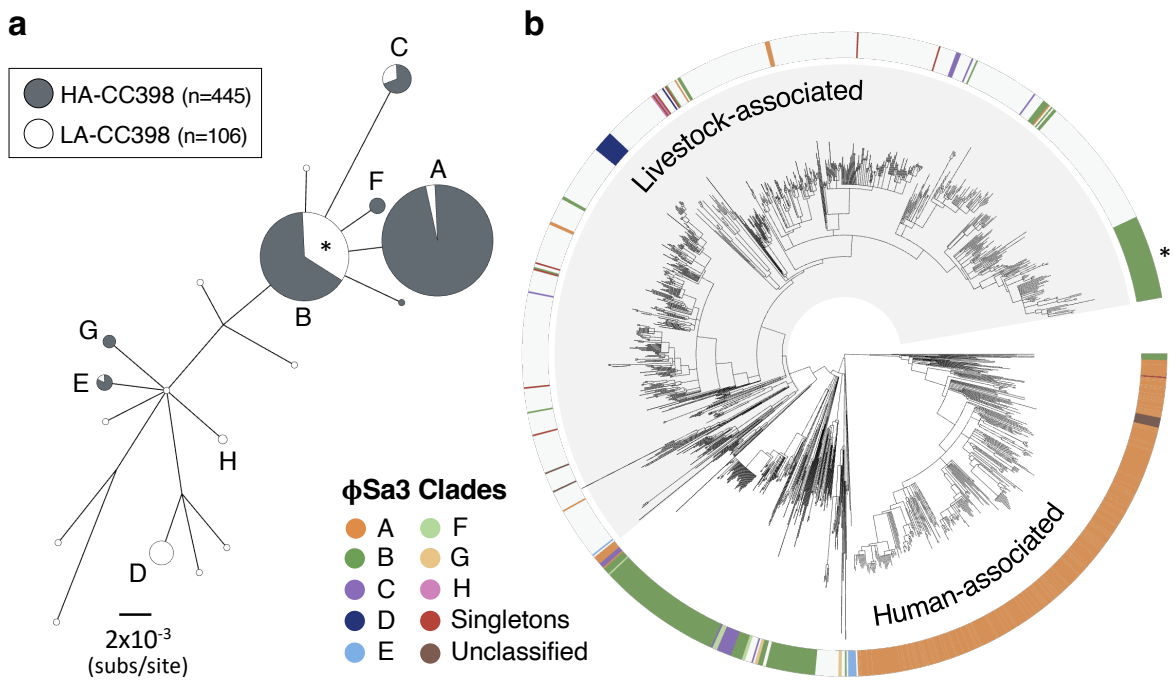


Figure 4. ϕSa3 prophages have been lost and acquired multiple times in both human-associated and livestock-associated CC398. (a) A maximum likelihood phylogeny based on 12 genes shared across the ϕSa3 prophages in our collection, with both low-support nodes (<70% bootstrap support) and branches <0.0018 subs/site collapsed. The latter cut-off is a conservative estimate of the maximum distance that could reflect divergence within CC398. It is the maximum pairwise distance between isolates carrying ϕSa3 prophages across 1,000 estimates from random samples of a core gene alignment of the same number of sites as is in our ϕSa3 prophage alignment. Node size correlates with the number of elements on a log-scale. Elements carried by isolates from human-associated CC398 (grey) and livestock-associated CC398 (white) isolates are indicated, and nodes that include multiple elements labelled (A-E). (b) These clades annotated on the CC398 phylogeny as an external ring. The element carried by the poultry-associated subclade of livestock-associated CC398 is indicated by *.

While ϕSa3 prophages are rare in livestock-associated CC398 (68/704 isolates), diversity within these elements indicate at least 15 (but likely more) acquisitions of these MGEs by livestock-associated CC398. The majority of these elements (69%) do not share a recent common ancestor with those in human-associated CC398. ϕSa3 prophages present in

livestock-associated CC398 generally show evidence of recent acquisition, with a notable exception being a previously described poultry-associated sub-clade (n=51) (Price et al. 2012; Larsen et al. 2016b; Pérez-Moreno et al. 2017; Tang et al. 2017). In our collection, 39% of the isolates in this clade are from poultry (compared to 4% across the rest of the livestock-associated clade). Low nucleotide diversity within the type B ϕSa3 consistently present in isolates within this clade suggests that it has been maintained since its acquisition approximately 21 years ago (95% CI: 1997-2001; Figure 1b and Figure 4-figure supplement 1).

Distinct route to multi-drug resistance in livestock-associated CC398

Livestock-associated CC398 is frequently multi-drug resistant. We see evidence of this in our data set where 81% of livestock-associated CC398 isolates carry one or more resistance genes for antibiotic classes other than tetracyclines and β-lactams (Figure 5-figure supplement 1, Figure 5-figure supplement 2, Supplementary File 7). 67% of livestock-associated CC398 isolates have genes associated with trimethoprim resistance (*dfrA*, *dfrK* or *dfrG*), 42% have genes associated with macrolide resistance (*ermA*, *ermB*, *ermC* or *ermT*), and 26% have genes associated with aminoglycoside resistance (*aadA, aphA* or *aphD*). Not only are resistance genes less common in human-associated CC398 (only 20% of isolates carry tetracycline resistance genes, and 9% carry trimethoprim resistance genes), they also differ in their relative frequencies. In particular, human-associated CC398 isolates more commonly carry genes associated with macrolide resistance (91%).

Spillover of livestock-associated CC398 into humans is associated with the acquisition of ϕSa3 prophages, but not a loss of resistance genes

ϕSa3 prophages are more common in human isolates (23%) than in livestock or companion animal isolates (11%) in our collection of livestock-associated CC398. To determine the significance of this association, we identified 70 phylogenetically independent clades that include isolates from both human and livestock or companion animal hosts (Figure 5-figure supplement 3). Comparisons across these groups revealed that isolates from humans are consistently more likely to carry an ϕSa3 prophage (McNemar's Chi-squared test: $p = 1.50 \times 10^{-3}$; Figure 5).

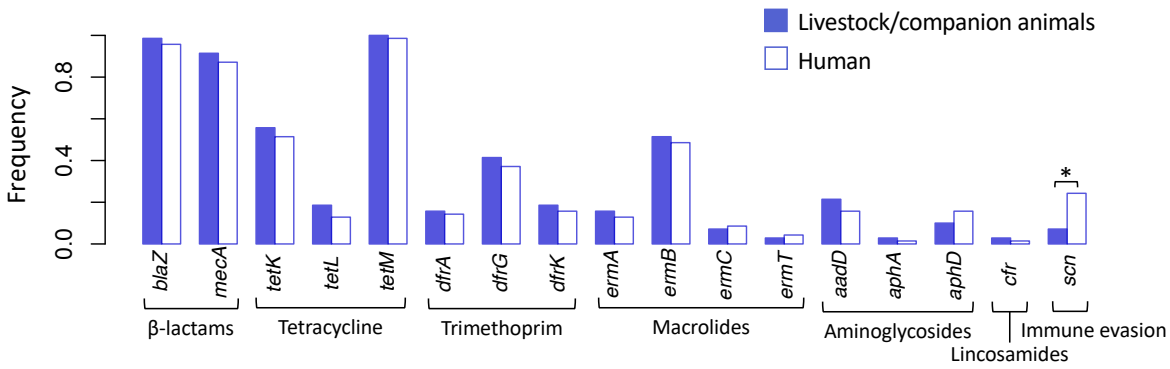


Figure 5. **Spillover of livestock-associated CC398 into humans is associated with acquisition of human immune-evasion genes.** 70 phylogenetically independent clades that include isolates from both humans and other species were identified within the livestock-associated clade. The plot shows the frequency with which these genes were identified within isolates from humans (right, empty bars) and non-human species (left, filled bars) in these groups. An asterisk indicates a significant difference based on McNemar's Chi-squared test ($p = 1.50 \times 10^{-3}$). No resistance genes differed significantly in their frequency across the human and non-human hosts ($p > 0.1$). The *scn* gene is always present in the human immune evasion cluster carried by ϕSa3 prophages, and therefore represents the presence of this element.

ϕSa3 prophages are also less common in isolates from pigs than from other non-human species (only 3% of pig isolates carry one). In 42/70 of our phylogenetically independent groups, the only non-human species was a pig, and none of the pig isolates from these groups carried an ϕSa3 prophage (while 14% of the human isolates did). In the remaining 28 groups (that included isolates from cows, horses and poultry), ϕSa3 prophages were observed in non-human hosts in 17% of groups, but still at a higher frequency in humans (39% of groups) (McNemar's Chi-squared test: $p = 0.04$).

In contrast, we found no evidence that the spillover of livestock-associated CC398 into humans is associated with the loss (or gain) of individual antibiotic resistance genes (Figure 5). This contrasts with the conclusions of a previous study that suggested that the resistance genes carried by livestock-associated CC398 are likely to be lost in human hosts (Sieber et al. 2019). We find that no resistance gene was significantly more or less common in humans than in livestock species (McNemar's Chi-squared tests: $p > 0.1$), nor was there a consistent shift in the overall number of resistance genes (there were more resistance genes in isolates from human hosts in 22/70 pairs and more in livestock hosts in 34/70 pairs).

Discussion

We have characterised the evolutionary dynamics of the three classes of MGEs that show the greatest changes in frequency across human- and livestock-associated CC398: the Tn*916* transposon and the SCC*mec* element, which are both common in livestock-associated CC398, and the ϕSa3 prophage, which is common in human-associated CC398. Despite a consistency in the relative frequencies of these elements across CC398, leading to their strong association with the transition to livestock, these three elements show a broad spectrum of dynamics. The Tn*916* transposon carrying *tetM* shows evidence of stable and consistent vertical transmission in livestock-associated CC398 and absence from human-associated CC398. The type Vc SCC*mec*, carrying not only *mecA*, but also *tetK* and *czrC*, has also been stably maintained by several lineages of livestock-associated CC398, but by a combination of vertical and horizontal transmission, and with occasional replacement with type IV elements. While type V SCC*mec* elements are also present in human-associated CC398, there is little evidence of their longer-term maintenance. Finally, while ϕSa3 prophages carrying a human immune-evasion gene cluster are rare in livestock-associated CC398 and common in human-associated CC398, there have been frequent gains and losses in both groups. These contrasting dynamics may reflect variation in the selective benefits and costs of the carriage of these MGEs by CC398, their availability in the environments encountered by CC398, and their mechanisms of transfer.

The three classes of MGEs whose dynamics we describe all employ different mechanisms of horizontal transfer, and this may both influence their intrinsic stability in the CC398 genome and their availability for acquisition from outside of CC398. Experimental studies have found that different types of MGEs vary in their rates of transfer between bacterial cells. In particular, in vitro rates of transfer of phage have been found to be several orders of magnitude higher than of transposons (Humphrey et al. 2021). A 16-day in vivo study of MGE dynamics for two strains of CC398 also observed variation in the mobility of MGEs: the Tn*916* transposon, the type V SCC*mec* and an ϕSa3 prophage were all stably maintained and not horizontally transferred, while other prophages and plasmids were both gained and lost by daughter cells (McCarthy et al. 2014). While variation in the intrinsic stability of MGEs in their host will influence their long-term dynamics, other factors may dominate. For instance, the long-term stability of Tn*916* within CC398 could emerge from a broad spectrum of short-term dynamics. It could result from a very low rate of loss, or frequent loss combined with a high selective cost of loss and a low probability of reacquisition from sources other

than very closely related cells. Therefore, the long-term dynamics we describe provide a unique insight into the relationship between these MGEs and CC398.

We find that a Tn*916* transposon that carries *tetM* was acquired once by CC398 and this coincided with its origin in European livestock in 1964 (95% CI: 1957-1970). Our observation of the loss of Tn*916* on terminal branches of the phylogeny suggests that while losses do occur, lineages that lose Tn*916* are either rapidly outcompeted by those that have maintained it, or they rapidly reacquire it from close relatives. Antibiotics, including tetracyclines, were first licensed for use as growth promoters in livestock in European countries in the 1950s, and were in common use by the end of that decade (Lowbury et al. 1958). While the use of antibiotics as growth promoters was banned by the European Union in 2006, tetracyclines remain the most commonly used antimicrobial class in livestock farming (WHO 2015; DANMAP 2019). Carriage of Tn*916* is therefore likely to have been associated with a strong selective benefit for livestock-associated CC398 ever since its emergence. However, as CC398's exposure to tetracyclines will be intermittent, the long-term stability of Tn*916* is also likely to reflect a low selective cost in the absence of treatment or a barrier to reacquisition following loss. Tn*916*-like elements are found across several bacterial genera (Clewell et al. 1995; Roberts and Mullany 2009), including other opportunistic pathogens in the respiratory microbiome of pigs (Holden et al. 2009; Hoa et al. 2011) and other *S. aureus* CCs (de Vries et al. 2009). This suggests that the stability of Tn*916* in CC398 is not due to the rarity of the element in the environments encountered by CC398, however it may be due to other barriers to successful transfer.

Previous studies have found evidence that the regulatory system of Tn*916* promotes its maintenance in a host through ensuring both that excision only occurs in the presence of tetracycline (or other transcription-limiting cell stress) and that any selective burden in the absence of treatment is minimised (Roberts and Mullany 2009). Nevertheless, Tn*916* has been found to have a much more dynamic association with lineages of other bacterial species (e.g. *Streptococcus pneumoniae* (D'Aeth et al. 2021)). The fitness costs of MGEs can be host (and even insertion locus) specific and can also be mitigated over time (Starikova et al. 2013; Durão et al. 2018), and therefore the stability of Tn*916* in livestock-associated CC398 might reflect a low cost that is specific to this lineage. A combination of high selective benefit, low selective cost and inaccessibility for reacquisition following loss could explain the remarkable stability of Tn*916* in livestock-associated CC398, and may in part explain the success of this lineage in livestock.

Our results indicate that the acquisition of an SCC*mec* occurred later in the expansion of livestock-associated CC398. The most common SCC*mec* in livestock-associated CC398, type Vc, carries *tetK* and *czrC* in addition to *mecA*. These additional resistance genes might be highly advantageous in livestock, and particularly in pigs. While all livestock-associated CC398 carry *tetM*, there is evidence that carrying *tetK* in addition to *tetM* is associated with increased fitness during exposure to sublethal concentrations of tetracycline (Larsen et al. 2016a). *czrC* is associated with heavy metal resistance (Cavaco et al. 2010), which is likely to be beneficial in the context of the common supplementation of animal feed with zinc oxide, which in pigs is commonly used to prevent diarrhoea in weaners (Nielsen et al. 2021). Additionally, *mecA* is likely to be beneficial because of the common use of beta-lactams in livestock farming, including third generation cephalosporins (Sjölund et al. 2016; Lekagul et al. 2019). The size of the type Vc element, combined with the replacements and truncations we observe suggests that it may come with a selective cost. In addition, SCC*mec* type Vc appears to be rare (at least in *S. aureus*). The element has been found in other staphylococci (*S. cohhii* in Vervet Monkeys (Hoefer et al. 2021)), but has not been reported in other *S. aureus* CCs.

Loss of the type Vc SCC*mec* on internal branches within livestock-associated CC398 is associated with replacement with a type IV SCC*mec* that only carries *mecA*. While we find evidence of only a single acquisition of the type Vc SCC*mec* by livestock-associated CC398, we find evidence of at least four acquisitions of type IV SCC*mec*. The more recent dates of acquisition, and an apparent association with livestock species other than pigs, might reflect a difference in selective pressures across different livestock species, or a loss of the type Vc element during transmission between livestock populations. The long-term maintenance of the type Vc SCC*mec* is consistent with a high selective benefit (particularly in pigs where this element is most common). However, its repeated loss and replacement with type IV elements is consistent with a low cost of this replacement, at least in certain contexts (perhaps in other livestock hosts), and might also reflect the rarity of the type V element relative to the type IV.

While the loss of the ϕSa3 prophage that carries a human immune-evasion gene cluster is associated with the transition to livestock, neither the loss nor the gain of these elements is likely to be a substantial hurdle for the adaptation of CC398 to human or non-human hosts. Their ubiquity in human-associated CC398 and frequent acquisition following transmission of

livestock-associated CC398 to humans (consistent with Sieber et al. 2019), is consistent with a strong benefit in the human host (Rohmer and Wolz 2021). On the other hand, as we find that these elements are frequently lost and generally absent in other host species, these elements may carry a selective burden outside of the human host. The diversity of ϕSa3 prophages in CC398 suggests they form a large pool of elements within human hosts (van Alen et al. 2018), likely reflecting their ubiquity in human carriage populations of *S. aureus* (Rohmer and Wolz 2021) and capacity to transfer between *S. aureus* lineages.

We observe one clear exception to the pattern of recent acquisitions of ϕSa3 in livestock-associated CC398 in response to spillover events: the acquisition of an ϕSa3 prophage at the base of a poultry-associated sub-clade of livestock-associated CC398 (Price et al. 2012; Larsen et al. 2016b; Pérez-Moreno et al. 2017; Tang et al. 2017). This clade was first described as a hybrid LA-MRSA CC9/CC398 lineage (Price et al. 2012), and was subsequently investigated as a lineage associated with human disease (Larsen et al. 2016b). The maintenance of an ϕSa3 in this lineage may reflect more frequent transmission via a human host, or an adaptation to poultry (a recent study suggested that ϕSa3 prophages might aid immune evasion in species other than humans (Jung et al. 2017)). Either way, this might make this lineage a greater immediate threat to public health.

While livestock-associated CC398 is found across a broad range of livestock species, it is most commonly associated with pigs. The dynamics of the SCC*mec* and ϕSa3 prophages that we have identified are both consistent with livestock-associated CC398 originating in pig farms and later spreading to other livestock species. The lower frequency of the type Vc SCC*mec* in species other than pigs could reflect random loss during transmission bottlenecks, or a reduced benefit of either *tetK* or *czrC* in these species. Similarly, the higher frequency of ϕSa3 prophages in other species might reflect an increased benefit or reduced cost, or more recent transmission via human hosts.

Our results reveal that LA-MRSA CC398 is a stably antibiotic-resistant pathogen that is capable of dynamic readaptation to humans. We find that Tn*916* and SCC*mec* are both stably maintained in livestock-associated CC398, across different livestock species and countries, and that neither of these MGEs (or other antibiotic resistance genes) tend to be lost when livestock-associated CC398 is transmitted to humans. This suggests that these MGEs are associated with a low selective cost in both livestock and human hosts, and therefore across variable levels and types of antibiotic exposure. The stability of these two

MGEs, combined with the capacity of livestock-associated CC398 to rapidly acquire the ϕSa3 prophage, underlines the threat posed by LA-MRSA public health. While SCC*mec* is less stably maintained than Tn*916*, our identification of several independent acquisitions of type IV SCC*mec* suggests both a strong selective benefit, not contingent on the carriage of *tetK* and *czrC*, and an availability of type IV elements. These dynamics predict that the impact of gradual reductions in antibiotic consumption on LA-MRSA are likely to be slow to be realised and that the forthcoming EU ban on medical zinc supplementation in pig feed may have a limited impact on LA-MRSA (European Medicines Agency 2017, 2020; DANMAP 2019). Further work is, however, required to understand the factors that underlie the acquisition and maintenance of resistance genes within LA-MRSA, and how they differ from human-associated MRSA lineages.

Methods

Data collection

All available genome sequence data relating to *S. aureus* CC398 was downloaded from public databases (www.ncbi.nlm.nih.gov/sra and www.ebi.ac.uk/ena; accessed 2019), with metadata in some cases obtained by request (Supplementary File 1). We additionally sequenced five isolates sampled from UK pig farms in 2018. All publicly available complete *S. aureus* genomes assemblies (www.ncbi.nlm.nih.gov; accessed 2019) were MLST typed using Pathogenwatch (https://pathogen.watch), and the 43 genomes identified as CC398 were added to our collection. After exclusion of low-quality assemblies and isolates mis-characterised as CC398, this led to a collection of 1,180 genomes.

Genomic library preparation and sequencing

For sequencing of the five UK pig farm isolates, genomic DNA was extracted from overnight cultures grown in TSB at 37°C with 200rpm shaking using the MasterPure Gram Positive DNA Purification Kit (Cambio, UK). Illumina library preparation and Hi-Seq sequencing were carried out as previously described (Harrison et al. 2013).

Genome assembly

We used sequence data from all isolates to generate *de novo* assemblies with *Spades* v.3.12.0 (Bankevich et al. 2012). We removed adapters and low quality reads with *Cutadapt* v1.16 (Martin 2011) and *Sickle* v1.33(Joshi and Fass 2011), and screened for contamination using *FastQ Screen* (Wingett and Andrews 2018). Optimal k-mers were identified based on average read lengths for each genome. All assemblies were evaluated using *QUAST* v.5.0.1(Gurevich et al. 2013) and we mapped reads back to *de novo* assemblies to investigate polymorphism (indicative of mixed cultures) using *Bowtie2* v1.2.2 (Langmead and Salzberg 2012). Low quality genome assemblies were excluded from further analysis (i.e., N50 <10,000, contigs smaller than 1kb contributing to >15% of the total assembly length, total assembly length outside of the median sequence length +/- one standard deviation, or >1,500 polymorphic sites). We identified genomes mischaracterised as CC398 via two approaches and excluded them from further analysis. First, we identified sequence types (STs) with *MLST-check* (Page et al. 2016) and grouped into CCs using the *eBURST* algorithm with a single locus variant (Francisco et al. 2009). Second, we constructed a neighbour-joining tree based on a concatenated alignment of MLST genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi* and *yqiL*) for our collection and 13 additional reference

genomes from other CCs, using the *ape* package in *R* and a K80 substitution model (Paradis et al. 2004).

We generated reference-mapped assemblies with *Bowtie2* using the reference genome S0385 (*GenBank* accession no. AM990992). For reference genomes, we generated artificial FASTQ files with *ArtificialFastqGenerator* (Frampton and Houlston 2012). Average coverage and number of missing sites in these assemblies were used as an additional quality control measure; genomes with average coverage <50x or with >10% missing sites were excluded.

We identified recombination in the reference-mapped alignment using both *Gubbins* v2.3.1(Croucher et al. 2015) and *ClonalFrame* (Didelot and Wilson 2015). We masked all the recombinant sites identified from our alignment. We additionally masked a region of ~123 kb that was identified as horizontally acquired from an ST9 donor in a previous study (Price et al. 2012).

Genome annotation and identification of homologous genes

We annotated *de novo* assemblies with *Prokka* v2.8.2 (Seemann 2014) and identified orthologous genes with *Roary* (Page et al. 2015) using recommended parameter values. We created a core gene alignment with *Roary* and identified recombinant sites using *Gubbins*. We identified antibiotic resistance genes using the *Pathogenwatch* AMR prediction module (Wellcome Sanger Institute), which uses *BLASTn* (Altschul et al. 1990) with a cut-off of 75% coverage and 80-90% percent identity threshold (depending on the gene) against a *S. aureus* AMR database.

Phylogenetic analyses

We carried out phylogenetic reconstruction for the reference-mapped alignment with *RAxML* v8.2.4 using the *GTR+Γ* model and 1,000 bootstraps(Stamatakis 2014). Sites where >0.1% of genomes showed evidence of recombination or had missing data were excluded from the analysis. We constructed dated phylogenies using *BEAST* v1.10 with a *HKY+Γ* model, a strict molecular clock, and constant population size coalescent prior, from the coding regions of the reference-mapped alignment (Drummond et al. 2012). We fit separate substitution models and molecular clocks to 1st/2nd and 3rd codon positions to reflect differences in selective constraint. We constructed phylogenies for three random subsamples of 250 isolates (200 from livestock-associated CC398 and 50 from human-associated CC398). Subsamples were non-overlapping except for 30 genomes representing the most divergent lineages within the

livestock-associated clade, to ensure a consistent description of the origin of this clade. We constructed additional phylogenies from subsamples that included only isolates from the LA-clade to establish that consistent rate estimates were returned over different evolutionary depths. We investigated temporal signal in each data set through a regression of root-to-tip distance against sampling date, and a permutation test of dates over tips (with clustering used to correct for any confounding of temporal and genetic structures) (Murray et al. 2016; Rambaut et al. 2016). Trees are visualized and annotated using *ITOL* (Letunic and Bork 2021).

Comparative analyses of mobile genetic elements

Genes associated with the transition to livestock were identified through comparing frequencies of homologous genes identified with *Roary* across human- and livestock-associated CC398. We investigated the association between these genes and MGEs through analysis of (i) physical locations within our *de novo* assemblies and the reference genomes, (ii) correlations in their presence/absence across our collection, and (iii) comparison with descriptions in the literature and in online databases of particular MGEs.

We confirmed the identity of the Tn*916* transposon by comparison with published descriptions and publicly available annotated sequences. SCC*mec* elements were initially categorised into types using a *BLASTn* search of all representative SCC*mec* types from the *SCCmecFinder* database in our *de novo* assemblies (Supplementary File 5) (Hanssen and Ericson Sollid 2006). ϕSa3 prophages were initially identified and categorised through identification of functional genes associated with the human immune evasion gene cluster. To further characterise diversity within these elements we identified genes associated with each element across our collection. We constructed alignments of genes within MGEs using *Clustal Omega* v1.2.3 (Sievers and Higgins 2018) and checked for misalignment by eye.

We analysed variation in both gene content and nucleotide diversity within shared genes for each MGE. We estimated pairwise nucleotide distances in concatenated alignments of shared genes for each type of element using the *ape* package in *R* (Paradis et al. 2004), and constructed maximum likelihood trees using *RaxML* and minimum spanning trees using *GrapeTree* (Zhou et al. 2018) to investigate co-phylogeny. We generated confidence intervals for our estimates of mean pairwise nucleotide distances within MGEs by re-estimating the mean distance from 1,000 bootstrapped samples of sites. We compared this to sites in the core genome by sampling the same number of sites as were in the MGE alignment from a concatenated alignment of core genes (generated by *Roary*). For SCC*mec*

we used ancestral state reconstruction in *BEAST* to infer the evolutionary dynamics of these elements and date their origins within CC398. This involved fitting a discrete traits model to the posterior distributions of trees, with each state representing a version of the element that had been independently acquired by CC398. We used a strict clock model that allowed for asymmetric rates of transitions between states, but we found that the results were robust to use of a symmetric model or a relaxed clock

Phylogenetically independent groups

To test the association between spillover into the human host and the presence of ϕSa3 prophages and antibiotic-resistance genes, we identified 70 phylogenetically independent clades of isolates that were sampled from both human and non-human hosts in the livestock-associated clade. We classed a gene as present in a host if it was observed in any of the isolates from that host in that clade.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The Illumina sequences generated in this study have been deposited in the NCBI short read archive under the accession numbers ERR3524650, ERR3524328, ERR3524354, ERR3524446 and ERR3524562. All other sequences used in this study are publicly available and their origins are described in Supplementary File 1.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

MM was funded by the Medical Research Council, co-funded by the Raymond and Beverly Sackler Fund. GGRM and LAW were supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (109385/Z/15/Z). GGRM was also supported by a ZELS BBSRC award (BB/L018934/1) and a Research Fellowship at Newnham College.

References

van Alen S., Ballhausen B., Kaspar U., Köck R., Becker K. 2018. Prevalence and genomic structure of bacteriophage phi3 in human-derived livestock-associated methicillin-resistant *Staphylococcus aureus* isolates from 2000 to 2015. Journal of Clinical Microbiology. 56.

van Alen S., Ballhausen B., Peters G., Friedrich A.W., Mellmann A., Köck R., Becker K. 2017. In the centre of an epidemic: Fifteen years of LA-MRSA CC398 at the University Hospital Münster. Veterinary Microbiology. 200:19–24.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403–410.

Anjum M.F., Marco-Jimenez F., Duncan D., Marín C., Smith R.P., Evans S.J. 2019. Livestock-Associated Methicillin-Resistant *Staphylococcus aureus* From Animals and Animal Products in the UK. Frontiers in Microbiology. 10.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A. v., Sirotkin A. v., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 19:455.

Brockhurst M.A., Harrison E., Hall J.P.J., Richards T., McNally A., MacLean C. 2019. The Ecology and Evolution of Pangenomes. Current Biology. 29:R1094–R1103.

Cavaco L.M., Hasman H., Stegger M., Andersen P.S., Skov R., Fluit A.C., Ito T., Aarestrup F.M. 2010. Cloning and occurrence of *czrC*, a gene conferring cadmium and zinc resistance in methicillin-resistant *Staphylococcus aureus* CC398 isolates. Antimicrobial Agents and Chemotherapy. 54:3605–3608.

Clewell D.B., Flannagan S.E., Jaworski D.D., Clewell D.B. 1995. Unconstrained bacterial promiscuity: the Tn*916*–Tn*1545* family of conjugative transposons. Trends in Microbiology. 3:229–236.

Croucher N.J., Page A.J., Connor T.R., Delaney A.J., Keane J.A., Bentley S.D., Parkhill J., Harris S.R. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research. 43:e15–e15.

D'Aeth J.C., van der Linden M.P.G., McGee L., de Lencastre H., Turner P., Song J.H., Lo S.W., Gladstone R.A., Sá-Leão R., Ko K.S., Hanage W.P., Breiman R.F., Beall B., Bentley S.D., Croucher N.J. 2021. The role of interspecies recombination in the evolution of antibiotic-resistant pneumococci. eLife. 10.

DANMAP. 2019. DANMAP: the Danish Programme for surveillance of antimicrobial consumption and resistance in bacteria from food animals, food and humans. 2019. .

Didelot X., Wilson D.J. 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLoS Computational Biology. 11.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 29:1969–1973.

Durão P., Balbontín R., Gordo I. 2018. Evolutionary Mechanisms Shaping the Maintenance of Antibiotic Resistance. Trends in Microbiology. 26:677–691.

European Medicines Agency. 2017. Zinc oxide. Available from https://www.ema.europa.eu/en/medicines/veterinary/referrals/zinc-oxide.

European Medicines Agency. 2020. European Surveillance of Veterinary Antimicrobial Consumption. Available from https://www.ema.europa.eu.

Frampton M., Houlston R. 2012. Generation of Artificial FASTQ Files to Evaluate the Performance of Next-Generation Sequencing Pipelines. PLOS ONE. 7:e49110.

Francisco A.P., Bugalho M., Ramirez M., Carriço J.A. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics 2009 10:1. 10:1–15.
Frost L.S., Leplae R., Summers A.O., Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology. 3:722–732.
Gladysheva I.P., Turner R.B., Sazonova I.Y., Liu L., Reed G.L. 2003. Coevolutionary patterns in plasminogen activation. Proceedings of the National Academy of Sciences. 100:9168–9172.
Gurevich A., Saveliev V., Vyahhi N., Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.
Haag A.F., Fitzgerald J.R., Penadés J.R. 2019. Staphylococcus aureus in Animals . Microbiology Spectrum. 7.
Hanssen A.-M., Ericson Sollid J.U. 2006. SCC*mec* in staphylococci: genes on the move. FEMS Immunology & Medical Microbiology. 46:8–20.
Harrison E.M., Paterson G.K., Holden M.T.G., Larsen J., Stegger M., Larsen A.R., Petersen A., Skov R.L., Christensen J.M., Zeuthen A.B., Heltberg O., Harris S.R., Zadoks R.N., Parkhill J., Peacock S.J., Holmes M.A. 2013. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. EMBO Molecular Medicine. 5:509–515.
He L., Zheng H.-X., Wang Y., Le K.Y., Liu Q., Shang J., Dai Y., Meng H., Wang X., Li T., Gao Q., Qin J., Lu H., Otto M., Li M. 2018. Detection and analysis of methicillin-resistant human-adapted sequence type 398 allows insight into community-associated methicillin-resistant Staphylococcus aureus evolution. Genome Medicine. 10.
Hoefer A., Boyen F., Beierschmitt A., Moodley A., Roberts M.C., Butaye P. 2021. Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus* from Vervet Monkeys (*Chlorocebus sabaeus*) in Saint Kitts. Antibiotics. 10.
Hsu L.-Y., Harris S.R., Chlebowicz M.A., Lindsay J.A., Koh T.-H., Krishnan P., Tan T.-Y., Hon P.-Y., Grubb W.B., Bentley S.D., Parkhill J., Peacock S.J., Holden M.T. 2015. Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. Genome Biology 2015 16:1. 16:1–13.
Humphrey S., Fillol-Salom A., Quiles-Puchalt N., Ibarra-Chávez R., Haag A.F., Chen J., Penadés J.R. 2021. Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. Nature Communications 2021 12:1. 12:1–13.
Islam M.Z., Espinosa-Gongora C., Damborg P., Sieber R.N., Munk R., Husted L., Moodley A., Skov R., Larsen J., Guardabassi L. 2017. Horses in Denmark Are a Reservoir of Diverse Clones of Methicillin-Resistant and -Susceptible *Staphylococcus aureus*. Frontiers in Microbiology. 0:543.
Jamrozy D., Coll F., Mather A.E., Harris S.R., Harrison E.M., MacGowan A., Karas A., Elston T., Estée Török M., Parkhill J., Peacock S.J. 2017. Evolution of mobile genetic element composition in an epidemic methicillin-resistant *Staphylococcus aureus*: temporal changes correlated with frequent loss and gain events. BMC Genomics 2017 18:1. 18:1–12.
Joshi N., Fass J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files . .
Jung P., Abdelbary M.M.H., Kraushaar B., Fetsch A., Geisel J., Herrmann M., Witte W., Cuny C., Bischoff M. 2017. Impact of bacteriophage Saint3 carriage on the immune evasion

capacity and hemolytic potential of *Staphylococcus aureus* CC398. Veterinary Microbiology. 200:46–51.
Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012 9:4. 9:357–359.
Larsen J., Clasen J., Hansen J.E., Paulander W., Petersen A., Larsen A.R., Frees D. 2016a. Copresence of tet(K) and tet(M) in livestock-associated methicillin-resistant *Staphylococcus aureus* clonal complex 398 is associated with increased fitness during exposure to sublethal concentrations of tetracycline. Antimicrobial Agents and Chemotherapy. 60:4401–4403.
Larsen J., Petersen A., Larsen A.R., Sieber R.N., Stegger M., Koch A., Aarestrup F.M., Price L.B., Skov R.L., Group for the D.M.S. 2017. Emergence of Livestock-Associated Methicillin-Resistant *Staphylococcus aureus* Bloodstream Infections in Denmark. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America. 65:1072.
Larsen J., Stegger M., Andersen P.S., Petersen A., Larsen A.R., Westh H., Agersø Y., Fetsch A., Kraushaar B., Käsbohrer A., Feβler A.T., Schwarz S., Cuny C., Witte W., Butaye P., Denis O., Haenni M., Madec J.-Y., Jouy E., Laurent F., Battisti A., Franco A., Alba P., Mammina C., Pantosti A., Monaco M., Wagenaar J.A., de Boer E., van Duijkeren E., Heck M., Domínguez L., Torres C., Zarazaga M., Price L.B., Skov R.L. 2016b. Evidence for Human Adaptation and Foodborne Transmission of Livestock-Associated Methicillin-Resistant *Staphylococcus aureus*. Clinical Infectious Diseases. 63:1349–1352.
Lekagul A., Tangcharoensathien V., Yeung S. 2019. Patterns of antibiotic use in global pig production: A systematic review. Veterinary and Animal Science. 7:100058.
Lekkerkerk W.S.N., van Wamel W.J.B., Snijders S. v., Willems R.J., van Duijkeren E., Broens E.M., Wagenaar J.A., Lindsay J.A., Vos M.C. 2015. What is the origin of livestock-associated methicillin-resistant *Staphylococcus aureus* clonal complex 398 isolates from humans without livestock contact? An epidemiological and genetic analysis. Journal of Clinical Microbiology. 53:1836–1841.
Letunic I., Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 49:W293–W296.
Li S., Skov R.L., Han X., Larsen A.R., Larsen J., Sørum M., Wulf M., Voss A., Hiramatsu K., Ito T. 2011. Novel Types of Staphylococcal Cassette Chromosome *mec* Elements Identified in Clonal Complex 398 Methicillin-Resistant *Staphylococcus aureus* Strains. Antimicrobial Agents and Chemotherapy. 55:3046.
Lowbury E.J.L., Selkon J.B., Smith H.W. 1958. Symposium on Epidemiological Risks of Antibiotics. Proceedings of the Royal Society of Medicine. 51:807.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.
Matuszewska M., Murray G.G.R., Harrison E.M., Holmes M.A., Weinert L.A. 2020. The Evolutionary Genomics of Host Specificity in Staphylococcus aureus. Trends in Microbiology. 28:465–477.
McCarthy A.J., Loeffler A., Witney A.A., Gould K.A., Lloyd D.H., Lindsay J.A. 2014. Extensive Horizontal Gene Transfer during Staphylococcus aureus Co-colonization In Vivo. Genome Biology and Evolution. 6:2697–2708.
Møller J.K., Larsen A.R., Østergaard C., Møller C.H., Kristensen M.A., Larsen J. 2019. International travel as source of a hospital outbreak with an unusual meticillin-resistant

Staphylococcus aureus clonal complex 398, Denmark, 2016. Eurosurveillance. 24:1800680.
Murray G.G.R., Wang F., Harrison E.M., Paterson G.K., Mather A.E., Harris S.R., Holmes M.A., Rambaut A., Welch J.J. 2016. The effect of genetic structure on molecular dating and tests for temporal signal. Methods in Ecology and Evolution. 7:80–89.
Nielsen C.L., Kongsted H., Sørensen J.T., Krogh M.A. 2021. Antibiotic and medical zinc oxide usage in Danish conventional and welfare-label pig herds in 2016–2018. Preventive Veterinary Medicine. 189:105283.
Page A.J., Cummins C.A., Hunt M., Wong V.K., Reuter S., Holden M.T.G., Fookes M., Falush D., Keane J.A., Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31:3691–3693.
Page A.J., Silva N. de, Hunt M., Quail M.A., Parkhill J., Harris S.R., Otto T.D., Keane J.A. 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microbial Genomics. 2:e000083.
Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.
Pérez-Moreno M.O., Centelles-Serrano M.J., Nogales-López J., Domenech-Spanedda M.F., Lozano C., Torres C. 2017. Unusual presence of the immune evasion gene cluster in livestock-associated MRSA of lineage CC398 causing peridural and psoas abscesses in a poultry farmer. Enfermedades Infecciosas y Microbiología Clínica. 35:651–654.
Postma B., Poppelier M.J., Galen J.C. van, Prossnitz E.R., Strijp J.A.G. van, Haas C.J.C. de, Kessel K.P.M. van. 2004. Chemotaxis Inhibitory Protein of *Staphylococcus aureus* Binds Specifically to the C5a and Formylated Peptide Receptor. The Journal of Immunology. 172:6994–7001.
Price L.B., Stegger M., Hasman H., Aziz M., Larsen J., Andersen P.S., Pearson T., Waters A.E., Foster J.T., Schupp J., Gillece J., Driebe E., Liu C.M., Springer B., Zdovc I., Battisti A., Franco A., Zmudzki J., Schwarz S., Butaye P., Jouy E., Pomba C., Porrero M.C., Ruimy R., Smith T.C., Robinson D.A., Weese J.S., Arriola C.S., Yu F., Laurent F., Keim P., Skov R., Aarestrup F.M. 2012. *Staphylococcus aureus* CC398: Host adaptation and emergence of methicillin resistance in livestock. mBio. 3:1–6.
Rambaut A., Lam T.T., Max Carvalho L., Pybus O.G. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution. 2.
Rankin D.J., Rocha E.P.C., Brown S.P. 2011. What traits are carried on mobile genetic elements, and why? Heredity 106:1. 106:1–10.
Roberts A.P., Mullany P. 2009. A modular master on the move: the Tn916 family of mobile genetic elements. Trends in Microbiology. 17:251–258.
Rohmer C., Wolz C. 2021. The Role of hlb-Converting Bacteriophages in Staphylococcus aureus Host Adaption. Microbial Physiology. 31:109–122.
Rooijakkers S.H.M., Ruyken M., Roos A., Daha M.R., Presanis J.S., Sim R.B., van Wamel W.J.B., van Kessel K.P.M., van Strijp J.A.G. 2005. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nature Immunology 2005 6:9. 6:920–927.
Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 30:2068–2069.
Sieber R.N., Larsen A.R., Urth T.R., Iversen S., Møller C.H., Skov R.L., Larsen J., Stegger M. 2019. Genome investigations show host adaptation and transmission of LA-MRSA

CC398 from pigs into Danish healthcare institutions. Scientific Reports 2019 9:1. 9:1–10.
Sieber R.N., Skov R.L., Nielsen J., Schulz J., Price L.B., Aarestrup F.M., Larsen A.R., Stegger M., Larsen J. 2018. Drivers and dynamics of methicillin-resistant livestock-associated staphylococcus aureus CC398 in pigs and humans in Denmark. mBio. 9.
Sievers F., Higgins D.G. 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 27:135–145.
Silva A.G. da, Baines S.L., Carter G.P., Heffernan H., French N.P., Ren X., Seemann T., Bulach D., Kwong J., Stinear T.P., Howden B.P., Williamson D.A. 2017. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant *Staphylococcus aureus*. Microbial Genomics. 3.
Sjölund M., Postma M., Collineau L., Lösken S., Backhans A., Belloc C., Emanuelson U., Beilage E.G., Stärk K., Dewulf J., Beilage E.G., GrosseLiesner B., Körk C.A., Lindberg A., Seemer H., Visschers V. 2016. Quantitative and qualitative antimicrobial usage patterns in farrow-to-finish pig herds in Belgium, France, Germany and Sweden. Preventive Veterinary Medicine. 130:41–50.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.
Starikova I., Al-Haroni M., Werner G., Roberts A.P., Sørum V., Nielsen K.M., Johnsen P.J. 2013. Fitness costs of various mobile genetic elements in Enterococcus faecium and Enterococcus faecalis. Journal of Antimicrobial Chemotherapy. 68:2755–2765.
Tang Y., Larsen J., Kjeldgaard J., Andersen P.S., Skov R., Ingmer H. 2017. Methicillin-resistant and -susceptible *Staphylococcus aureus* from retail meat in Denmark. International Journal of Food Microbiology. 249:72–76.
Thammavongsa V., Kim H.K., Missiakas D., Schneewind O. 2015. Staphylococcal manipulation of host immune responses. Nature Reviews Microbiology 2015 13:9. 13:529–543.
Turner N.A., Sharma-Kuinkel B.K., Maskarinec S.A., Eichenberger E.M., Shah P.P., Carugati M., Holland T.L., Fowler V.G. 2019. Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. Nature Reviews Microbiology 2019 17:4. 17:203–218.
Vandendriessche S., Vanderhaeghen W., Larsen J., de Mendonça R., Hallin M., Butaye P., Hermans K., Haesebrouck F., Denis O. 2014. High genetic diversity of methicillin-susceptible *Staphylococcus aureus* (MSSA) from humans and animals on livestock farms and presence of SCCmec remnant DNA in MSSA CC398. Journal of Antimicrobial Chemotherapy. 69:355–362.
de Vries L.E., Christensen H., Skov R.L., Aarestrup F.M., Agersø Y. 2009. Diversity of the tetracycline resistance gene tet(M) and identification of Tn916- and Tn5801-like (Tn6014) transposons in Staphylococcus aureus from humans and animals. Journal of Antimicrobial Chemotherapy. 64:490–500.
van Wamel W.J.B., Rooijakkers S.H.M., Ruyken M., van Kessel K.P.M., van Strijp J.A.G. 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on β-hemolysin-converting bacteriophages. Journal of Bacteriology. 188:1310–1315.
Ward M.J., Gibbons C.L., McAdam P.R., van Bunnik B.A.D., Girvan E.K., Edwards G.F., Fitzgerald J.R., Woolhouse M.E.J. 2014. Time-scaled evolutionary analysis of the

transmission and antibiotic resistance dynamics of *Staphylococcus aureus* clonal complex 398. Applied and Environmental Microbiology. 80:7275–7282.
WHO. 2015. Global Action Plan on Antimicrobial Resistance. .
Wingett S.W., Andrews S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research. 7:1338.
Zhou Z., Alikhan N.-F., Sergeant M.J., Luhmann N., Vaz C., Francisco A.P., Carriço J.A., Achtman M. 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Research. 28:1395.
Zou G., Matuszewska M., Jia M., Zhou J., Ba X., Duan J., Zhang C., Zhao J., Tao M., Fan J., Zhang X., Jin W., Cui T., Zeng X., Jia M., Qian X., Huang C., Zhuo W., Yao Z., Zhang L., Li S., Li L., Huang Q., Wu B., Chen H., Tucker A.W., Grant A.J., Holmes M.A., Zhou R. 2022. A Survey of Chinese Pig Farms and Human Healthcare Isolates Reveals Separate Human and Animal Methicillin-Resistant Staphylococcus aureus Populations. Advanced Science. 9.

Supplementary Figure Legends

Figure 1-figure supplement 1. The temporal, host-species and geographic distribution of our collection of CC398 isolates. (a) Phylogeny of CC398 with rings showing the host species and countries of origin for each isolate (groups with n<10 not shown). The blue outline indicates the livestock-associated clade, and the red outline indicates the 5 most recent isolates in our collection (sampled in 2018), from pigs on UK farms. (b), (c) and (d) show the variation in sampling date, host species and country across the livestock-associated (blue, lower) and human-associated (red, upper) clades.

Figure 1-figure supplement 2. Different outgroups consistently identify the root of CC398 within human-associated CC398. We constructed maximum likelihood phylogenies using a reference-mapped alignment of CC398 combined with four outgroups from sequence types (STs) 291, 30, 97 and 5; and using a core genome alignment of our *de novo* assemblies and a midpoint rooting. The outgroups we used covered a range of distances from the base of CC398: ST291 (ERR2729529) is ~0.005 subs/site, ST30 (ERS1420125) is ~0.012 subs/site, and ST97 and ST5 (ERR2729579 and SRS613151) are ~0.015 subs/site. The reference-mapped phylogenies that were rooted using ST291, ST30 and ST97, and the midpoint-rooted core genome phylogeny all showed a consistent root, which is shown in the figure (and indicated by 1). A different root was obtained when we used the outgroup of ST5 (location indicated by 2). This root is on a neighbouring branch to the root found in the 4 other reconstructions, and results in CC398 being rooted on the branch leading to a single isolate (ZTA09_03734_9HSA). Livestock-associated CC398 is indicated by a blue box with grey shading.

Figure 1-figure supplement 3. A consistent estimate of the age of the livestock-associated clade. Results of BEAST dating analyses estimating (a) the origin of a shallower subclade within the livestock-associated clade, (b) the origin of the entire livestock-associated clade, and (c) the origin of CC398. (a) The figure shows a schematic representation of the CC398 phylogeny indicating the nodes of interest (A-C), and our sampling strategy. We randomly sampled our data set three times to generate samples of 250 isolates (200 from the livestock-associated clade and 50 from the human-associated clade). Samples overlapped by only 30 isolates that represent the most divergent lineages of the livestock-associated clade, to ensure a consistent description of the most recent common ancestor. Each sample had the same range of sampling dates (1993-2018). (a) also describes the results of a regression of root-to-tip distance against sampling date (correlation coefficient and estimate of the date of the most recent common ancestor) for each sample (1-3), and subsamples that include isolates from (A) only the main livestock-associated clade, (B) only the livestock-associated clade, and (C) the entire sample. As we observed stronger

temporal signal for A and B, than for C, we estimated dated trees using BEAST for each of these 9 subsampled data sets. We observed consistent estimates of evolutionary rate across all these analyses (b). Rates at 1st/2nd codon positions are shown as circular points, and at 3rd codon positions as square points). These analyses also returned broadly consistent estimates of dates (c). Although we found that estimates from subsamples that included outgroups of the node being dated returned more precise and marginally more recent estimates of age, likely due to more information about the location of the root.

Figure 1-figure supplement 4. Evidence of temporal signal is present across in our subsampled datasets, but is stronger when isolates from the human-associated group are excluded. Regressions of root-to-tip distance against sampling date for each of our data sets, rooted to minimise residual mean squares. (a)-(c) show the results for samples 1, 2 and 3 for clade a (described in Figure 1-figure supplement 3), (d)-(f) the results for samples 1, 2 and 3 for clade b, and (g)-(i) samples 1, 2 and 3 for clade c. For all data sets a randomisation test indicated that these correlations were unlikely to have arisen by change ($p<0.01$). Correlation coefficients (r) and estimates of the time of the most recent common ancestor (tMRCA) based on the regression are shown for each data set.

Figure 1-figure supplement 5. Livestock-associated and human-associated CC398 have divergent accessory genomes, and genes whose presence/absence most clearly distinguish these groups (except one) are associated with a Tn*916* transposon, SCC*mec* and ϕSa3 prophages. (a) A plot of the 1st and 2nd principal components of accessory genome content, with isolates from human-associated CC398 in red and isolates from livestock-associated CC398 in blue. This was constructed using the package *adegenet* in *R* [1]. (b) Comparison of gene frequencies across the human-associated and livestock-associated groups. Genes present in <20% of human-associated CC398 and >80% of livestock-associated CC398 and genes presence in >80% of human-associated CC398 and <20% livestock-associated CC398 are highlighted, with genes associated with the Tn*916* transposon shown in purple (all are overlapping as they have identical frequencies), genes associated with SCC*mec* shown in turquoise, and genes associated with ϕSa3 prophages shown in blue. The one gene that distinguishes livestock-associated CC398 from human-associated CC398 that isn't associated with one of these three MGEs is shown as a red circle (*tatC*).

Figure 2-figure supplement 1. Evidence of repeated excision of Tn*916* transposon in the livestock-associated clade. (a) Purple points indicate which 5 isolates lack the Tn*916* transposon in the livestock-associated clade. (b) Top: A gene map showing the Tn*916* transposon from the 1_1429 reference strain and the two flanking genes (dark grey). Bottom: The integration site in the

62951 strain within which the transposon is absent. The percentage nucleotide identity between the two flanking genes is indicated below the bars connecting the top and bottom gene maps.

Figure 2-figure supplement 2. Type V and IV SCC*mec* elements identified in CC398 through a *BLASTn* search or representative types. Initial identification of SCC*mec* types was carried out by *BLASTn* search of all SCC*mec* reference sequences on the *SCCmecFinder* extended database. As the SCC*mec* element in our dataset was commonly separated onto multiple contigs, we found that contiguous hits of the entire element were rare. Therefore, to determine the presence of a particular SCC*mec* type we considered the combined length of high-identity hits (hits that have >95% nucleotide identity and are ≥5% of the length of the element). We categorised elements into SCC*mec* types based on the overall length of the matched region, and described strains with best match lengths of <50% as unknown. The CC398 phylogeny is annotated with the results of this analysis. The innermost ring shows the presence/absence of *mecA*, and outer rings show the percentage length match for each type/sub-type (Supplementary File 7).

Figure 3-figure supplement 1. Most of the shorter versions of the type Vc SCC*mec* element in CC398 can be attributed to deletion events. Of the 540 isolates that carry a type V SCC*mec* in CC398, 204 have >2 genes absent. There are 3 common forms of gene absence (B, C and D). (a) Gene maps showing common patterns of gene absence. (b) A minimum spanning tree of a concatenated alignment of the 40 genes associated with the full-length SCC*mec* type V. Points represent groups of elements that differ by a maximum of 1 SNP, with point size correlated with group size on a log-scale, and colours representing categories described in (a). (c) Annotation of the full-length and truncated categories onto the core genome phylogeny. (d) Histograms showing pairwise distances (per shared nucleotide site) between individual elements in each group and the least divergent member of Group A. Histograms are shown on a log(x+1) scale due to the high frequency of low divergence elements. The dashed vertical lines show the average pairwise distance between isolates within the livestock-associated clade based on a core gene alignment, and the dotted lines show 10-times this value. Groups A, B and C are only found in the livestock-associated clade. Due to the low level of diversity in group D, versions from the human-associated and livestock-associated clade are shown together, while for group E they are shown separately (HA/LA). There are 13 isolates from the human-associated clade in group E. Of these 13 isolates, only 3 show sufficiently low divergence to be consistent with a recent common ancestor with the majority of elements from Group A, within CC398 (for these three elements divergence is $<1.7 \times 10^{-4}$ /site). The three isolates carrying these elements all fall within the recent Danish hospital outbreak clade, and show 100% identity with the majority of the elements within this clade (that fall within group D).

Figure 3-figure supplement 2. There have been at least four independent acquisitions of type IV SCC*mec* within livestock-associated CC398. (a) An minimum-spanning tree based on a concatenated alignment of the 12 genes shared across all type IV SCC*mec* (n=148 isolates). Clades are distinguished based on nucleotide distances within these shared genes and by differences in the genes present in these elements. Clade A is IVa(1), Clade B is IVc, Clade C is IVa(2), and Clade D is IVa(3). IVa(2) and IVa(3) only differ by one SNP in the alignment of 12 shared genes, but they have different gene contents, suggesting divergent origins. (b) Clades of type IV elements mapped onto the core genome phylogeny.

Figure 3-figure supplement 3. Tn*916* was acquired before the current complement of SCC*mec* elements in livestock-associated CC398, SCC*mec* type V was acquired before SCC*mec* type IV, and SCC*mec* type V was acquired by livestock-associated CC398 before human-associated CC398. Date estimates based on ancestral state reconstructions with BEAST over dated phylogenies inferred from three subsampled data sets. (A) A dated phylogeny inferred from one subsample (no. 1) with branches coloured by inferred SCC*mec* element at their descendent node (>95% posterior support), and earliest nodes (>95% posterior support) for each element labelled. (B) Date estimates for the earliest nodes at which each SCC*mec* type is inferred to be present (>95% posterior support) across the three data sets. Points are median values and bars are 95% confidence intervals. Estimates are shown for three versions of SCC*mec* type IVa that were distinguished based on the divergence within this element (see Figure 3-figure supplement 3). Only one estimate is provided for SCC*mec*-IVa(3) because it is only observed in one of the subsampled data sets.

Figure 4-figure supplement 1. Maintenance of a ϕSa3 prophage in a poultry-associated clade of livestock-associated CC398 for around 21 years. (a) A minimum spanning tree of the type B ϕSa3 prophage present in the poultry-associated clade of livestock-associated CC398. 55 isolates in our collection of CC398 carry a version of this element. We identified 34 genes shared across these 55 versions of the element. After pruning genes with homologues, and difficult to align regions 29/34 genes were used to construct the tree. Points represent groups of identical elements, with point size correlated with group size on a log-scale, and colours representing clades that differ from the basal group by >1 SNP. (b) These clades annotated onto the CC398 phylogeny as an external ring. 51/55 isolates carrying this element fall within the poultry-associated clade and 4 other isolates are from other clades within livestock-associated CC398 from human hosts. Within the avian clade we see little diversity (maximum pairwise distance is 1 SNP). The elements from outside of this clade show greater divergence, but this could still represent divergence within livestock-associated CC398.

Figure 5-figure supplement 1. Patterns in the presence/absence of antibiotic resistance genes vary across livestock-associated CC398 and human-associated CC398 in addition to the three genes associated with Tn*916* and SCC*mec*. Variation in presence of antibiotic resistance genes across CC398 mapped onto the core genome phylogeny. All genes that are present above a threshold of 5% are represented, with colours representing different antibiotic classes.

Figure 5-figure supplement 2. Differences in antibiotic resistance gene frequencies across human- and livestock-associated CC398. (a) Comparison of the frequencies of resistance genes across clades, with the livestock-associated group shown in the filled bar on the left and the human-associated group the empty bar on the right. (b) Comparison of only isolates carrying *mecA* from the two groups. (c) and (d) Venn diagrams that show the relationship between the presence of genes associated with different antibiotic classes in MRSA isolates from both groups.

Figure 5-figure supplement 3. Phylogeny of the livestock-associated clade showing locations of 70 phylogenetically independent groups of isolates from human (red) and livestock/companion animal (blue) hosts. The phylogeny is a maximum likelihood phylogeny of the livestock-associated clade, rooted using human-associated CC398, with nodes with <70% bootstrap support collapsed. 70 well-supported independent clades that contain both isolates from human and livestock/companion animal hosts were identified across the livestock-associated clade, and were used to test for differences in the frequency of genes associated with both antibiotic resistance and adaptation to the human host across these two groups. The small number of isolates from wild or domestic animals (other than livestock and horses) were excluded from the analysis.

Supplementary Table Legends

Supplementary File 1. Metadata for all isolates used in this study. Strain names, country of origin, source (host species), year, accession numbers and references for all isolates.

Supplementary File 2. Genes that most strongly distinguish human- and livestock-associated CC398, and their association with mobile genetic elements. Our gene identifiers and the gene locations and locus tags in published reference genomes are provided.

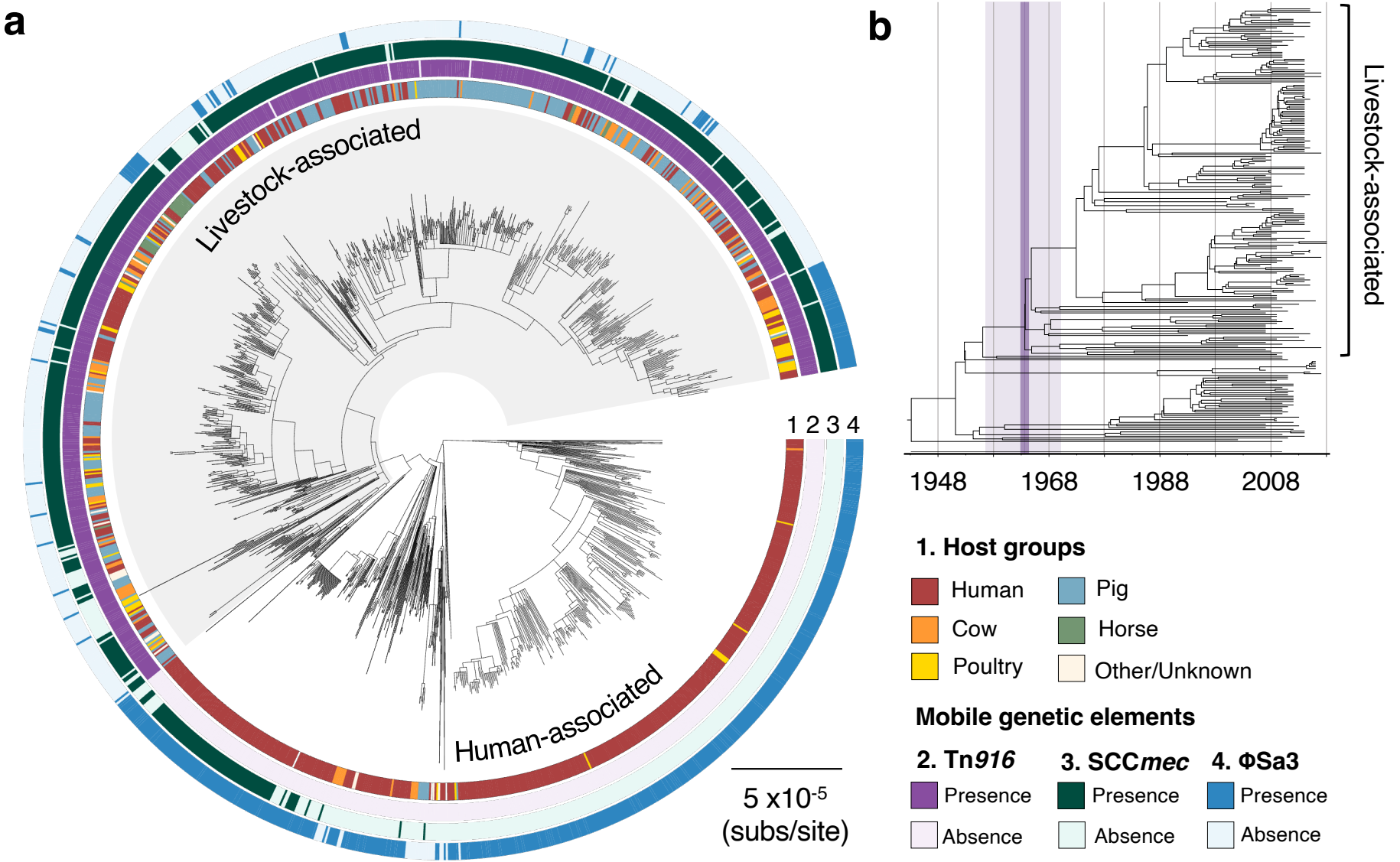
Supplementary File 3. Description of the presence of MGEs and annotation of MGE types and clades. The presence/absence of genes and MGEs are described by 1/0, and types and clades that are presented in the text are described.

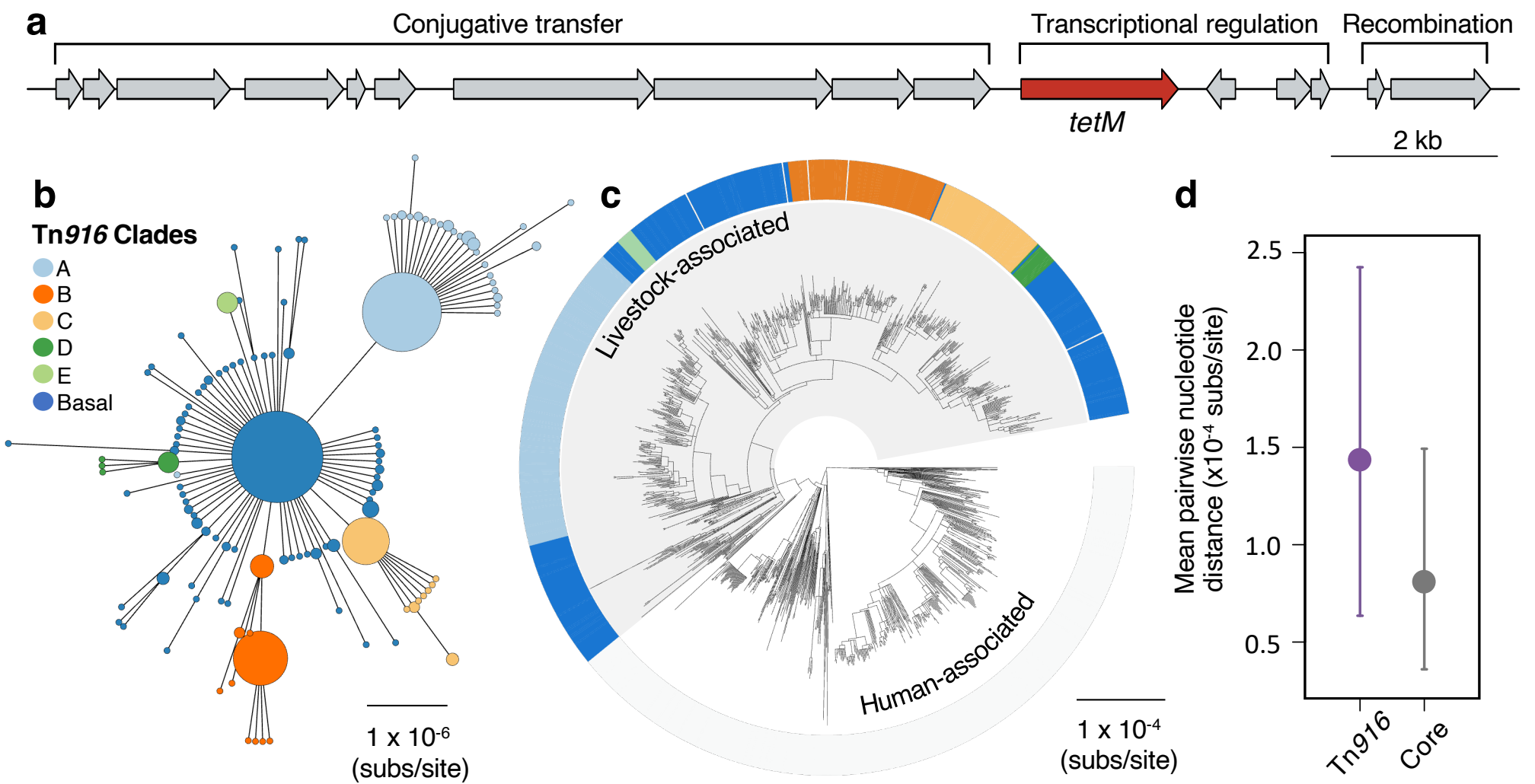
Supplementary File 4. Description of the genes in the Tn*916* element. The genes in the Tn*916* element used in our analyses are described in the reference genome 1_1439.

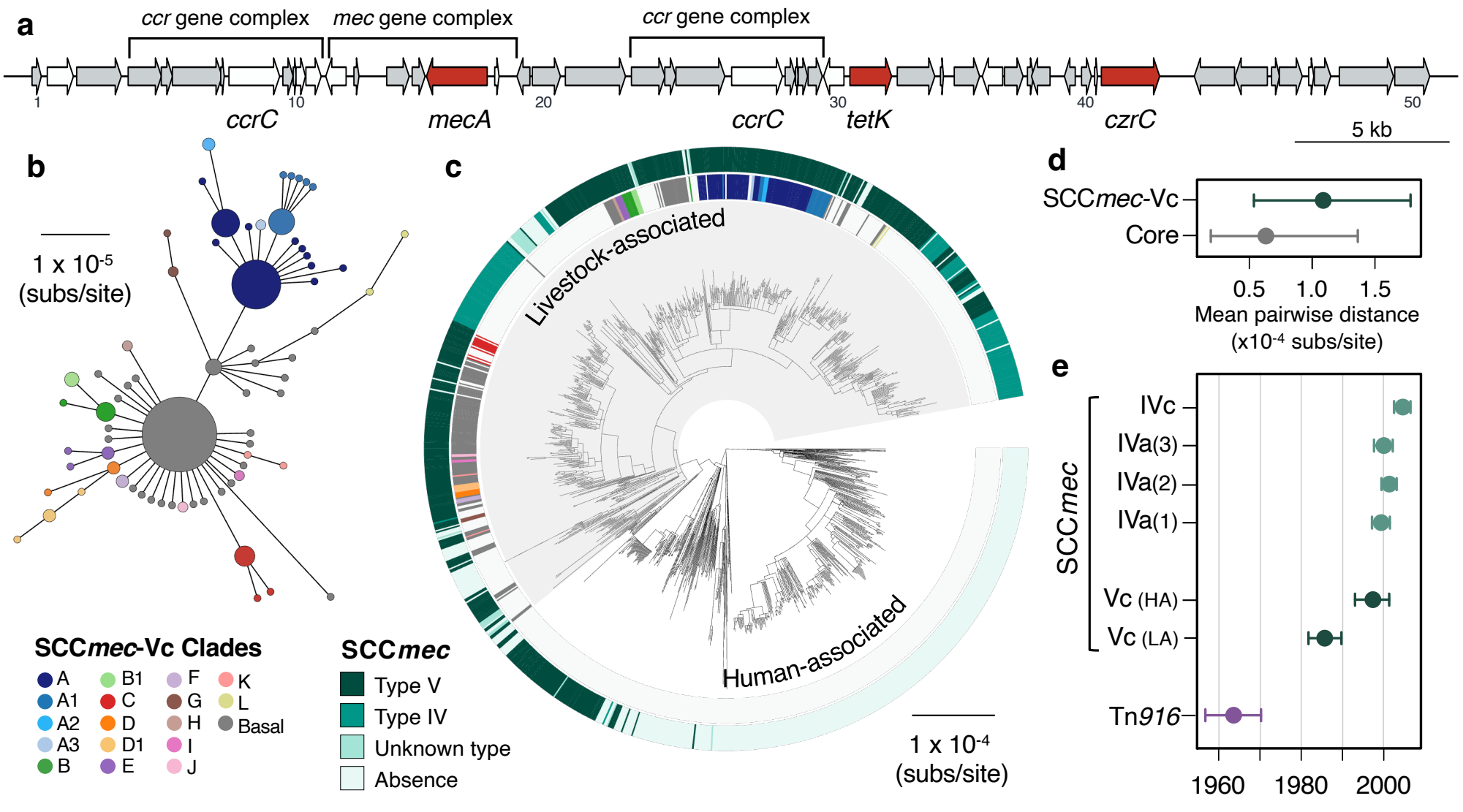
Supplementary File 5. Reference SCC*mec* elements used in BLAST typing.

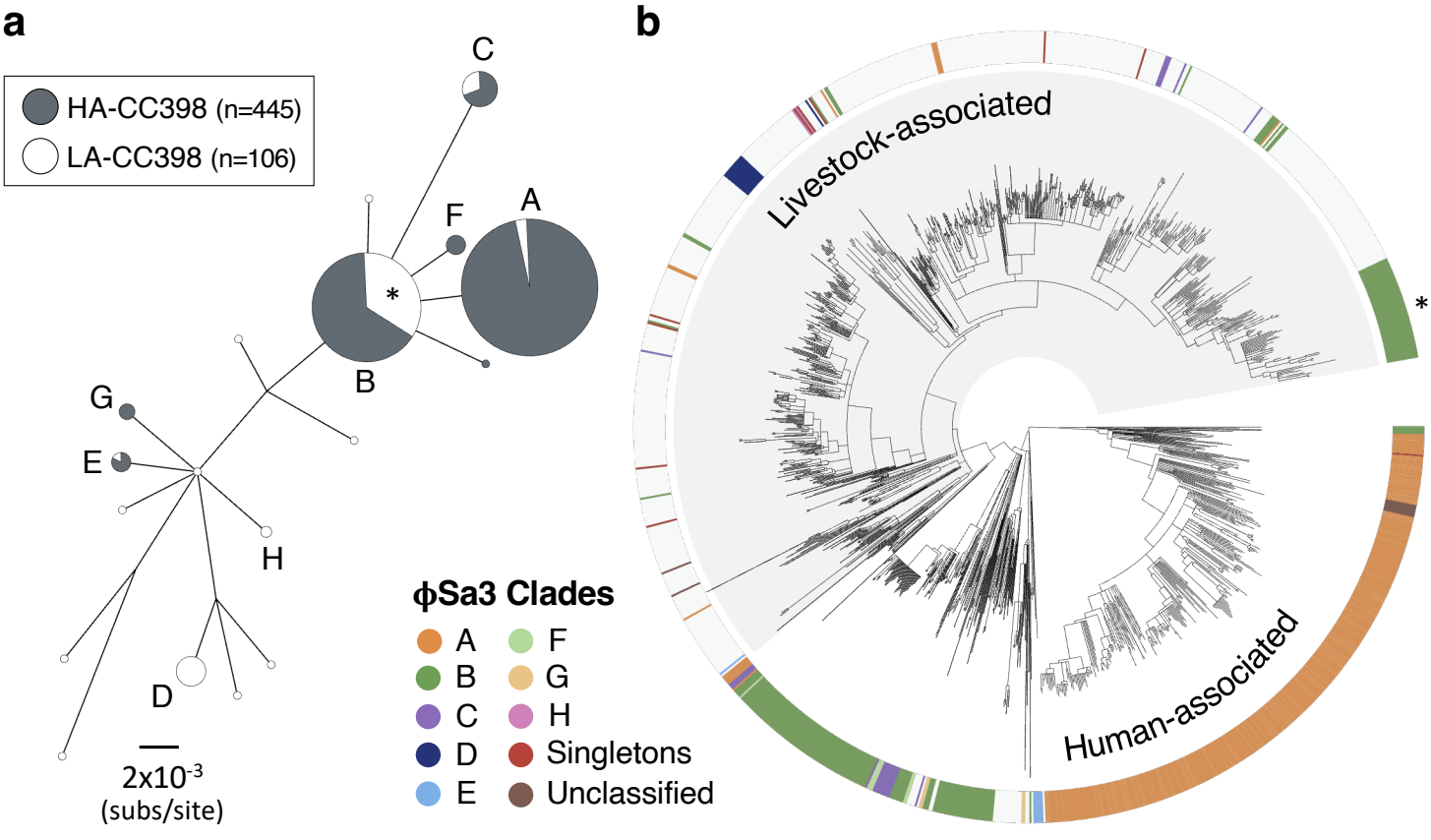
Supplementary File 6. Description of the genes in the type V SCC*mec* element. The genes in the SCC*mec* type Vc element used in our analyses are described in the reference genome 12_LA_293.

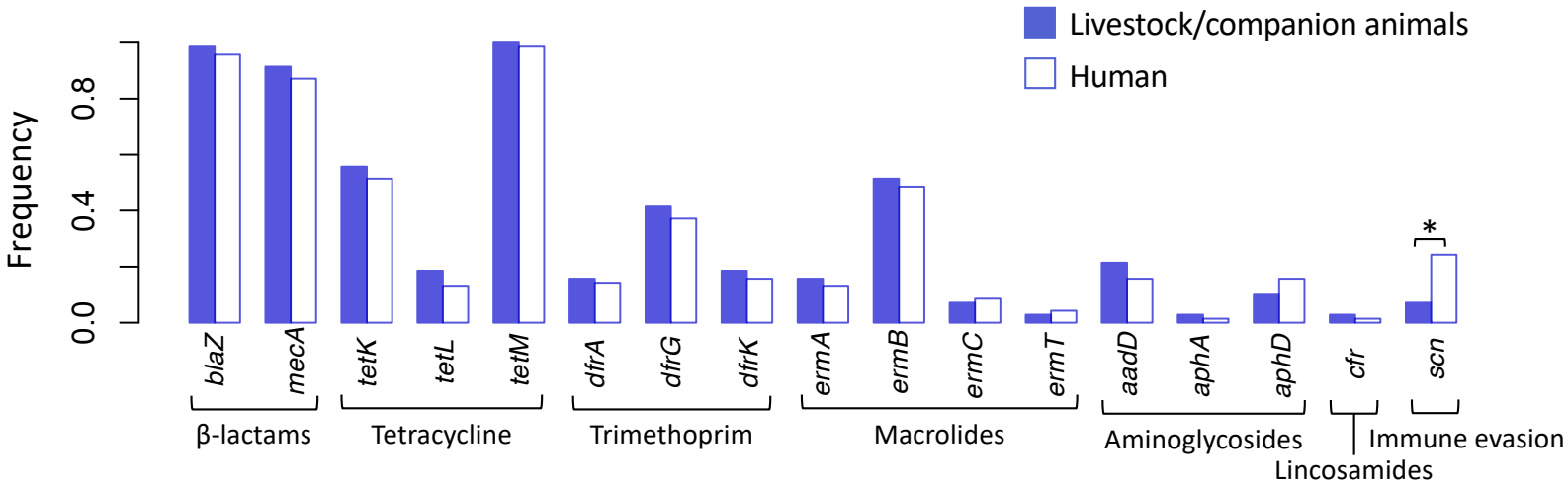
Supplementary File 7. AMR genes identified by PathogenWatch. Gene presence/absence is described by 1/0.


a
Livestock-associated
Human-associated
1 2 3 4
5 x10-5
(subs/site)
b
Livestock-associated
1948
1968
1988
2008
1. Host groups
Human
Pig
Cow
Horse
Poultry
Other/Unknown
Mobile genetic elements
2. Tn916
3. SCCmec
4. ΦSa3
Presence
Absence
Presence
Absence
Presence
Absence


a
Conjugative transfer
Transcriptional regulation
Recombination
tetM
2 kb
b
Tn916 Clades
A
B
C
D
E
Basal
1 x 10-6
(subs/site)
c
Livestock-associated
Human-associated
1 x 10-4
(subs/site)
d
Mean pairwise nucleotide distance (x10-4 subs/site)
2.5
2.0
1.5
1.0
0.5
Tn916
Core


a
ccr gene complex
mec gene complex
ccr gene complex
1
10
20
30
40
50
ccrC
mecA
ccrC
tetK
czrC
5 kb
b
1 x 10-5
(subs/site)
SCCmec-Vc Clades
A
A1
A2
A3
B
B1
C
D
D1
E
F
G
H
I
J
K
L
Basal
SCCmec
Type V
Type IV
Unknown type
Absence
c
Livestock-associated
Human-associated
1 x 10-4
(subs/site)
d
SCCmec-Vc
Core
0.5
1.0
1.5
Mean pairwise distance
(x10-4 subs/site)
e
SCCmec
IVc
IVa(3)
IVa(2)
IVa(1)
Vc (HA)
Vc (LA)
Tn916
1960
1980
2000


a
HA-CC398 (n=445)
LA-CC398 (n=106)
C
F
A
*
B
G
E
H
D
2x10-3
(subs/site)
ϕSa3 Clades
A
B
C
D
E
F
G
H
Singletons
Unclassified
b
Livestock-associated
*
Human-associated


Livestock/companion animals
Human
Frequency
0.0
0.4
0.8
blaZ
mecA
tetK
tetL
tetM
dfrA
dfrG
dfrK
ermA
ermB
ermC
ermT
aadD
aphA
aphD
cfr
scn
*
β-lactams
Tetracycline
Trimethoprim
Macrolides
Aminoglycosides
Lincosamides
Immune evasion