Half a century of stable antibiotic resistance in livestock-associated *Staphylococcus aureus* and its dynamic readaptation to humans

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.

Abstract

Mobile genetic elements (MGEs) often carry genes that benefit their bacterial hosts. In methicillin-resistant *Staphylococcus aureus* (MRSA), MGEs have been associated with antibiotic resistance, virulence, and host adaptation. Clonal-complex (CC) 398 is the dominant MRSA in European livestock, and a growing cause of human infections. To understand the risk posed by livestock-associated MRSA to human health, we have used a collection of 1,180 CC398 genomes, sampled from several livestock species and humans, with a broad geographic distribution and spanning 27 years, to reconstruct the dynamics of the MGEs. 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 vertically inherited for 57 years. This was followed by the acquisition of a large SCC*mec* type V element that carries methicillin, tetracycline and heavy metal resistance genes. This has been maintained within livestock-associated CC398 for at least 35 years, with occasional truncations and replacements with other, smaller type IV SCC*mec* elements. In contrast, a class of prophages that carry a human immune-evasion gene cluster, that are largely absent from livestock-associated CC398, have been repeatedly gained and lost across both human- and livestock-associated CC398. The variable dynamics of these three MGEs means that when livestock-associated MRSA infects humans, re-adaptation to the human host outpaces the loss of antibiotic resistance. Moreover, the stability of both Tn*916* and SCC*mec* suggests that they may persist despite ongoing reductions in antibiotic and zinc oxide use in farming.

Introduction

Mobile genetic elements (MGEs) play an important role in the evolution of bacterial pathogens. They can move rapidly between bacterial genomes and also be vertically inherited through stable integration into a host genome. As MGEs often carry genes associated with virulence and antibiotic resistance [1,2], 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 [3].

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 [4–6]. Most MRSA clonal complexes (CCs) show an association with specific MGEs, but the dynamics of these MGEs is not widely understood [5,7]. Livestock-associated (LA) MRSA CC398, has become the dominant MRSA in European livestock. Its rise is particularly evident in Danish pig farms where the proportion of MRSA-positive herds has increased from <5% in 2008 to 90% in 2018 [8,9], but it has also been observed in other European countries, and other livestock species [10–12]. The rise of LA-MRSA CC398 is of particular concern because it has been associated with increasing numbers of human infections, in both people with and without direct contact with livestock [8,13,14]. Intensification combined with high levels of antibiotic use in farming has led to concerns that livestock may serve as reservoirs of antibiotic-resistant human infections [15,16]. Understanding the emergence and success of CC398 in European livestock 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 [17–19]. They identified a largely methicillin-resistant livestock-associated clade of CC398, that falls as either sister to [18] or within [17,19] a largely methicillin-sensitive 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. However, little is known about the dynamics of the MGEs that carry these genes within CC398. Here, we undertake a comprehensive reconstruction of the dynamics of MGEs associated with livestock-associated CC398 in order to better understand the success of this lineage in livestock. We find that three MGEs show a strong association with the transition to livestock, and yet have 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 S1, Table S1).

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 has been 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, Figure S2) [17,19].

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 included 250 isolates. Our estimates of the evolutionary rate were consistent across all reconstructions (1.1-1.6 $\times 10^{-6}$ subs/site/year), and similar to estimates from previous studies of CC398 [18] and other *S. aureus* CCs [20]. This led to a consistent estimate of the origin of the livestock-associated clade of approximately 1964 (95% CI: 1957-1970) (Figure 1b, Figures S3 and S4).

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, IEC) [17]. 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 IEC (Figures 1a, S5, Table S2). 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.

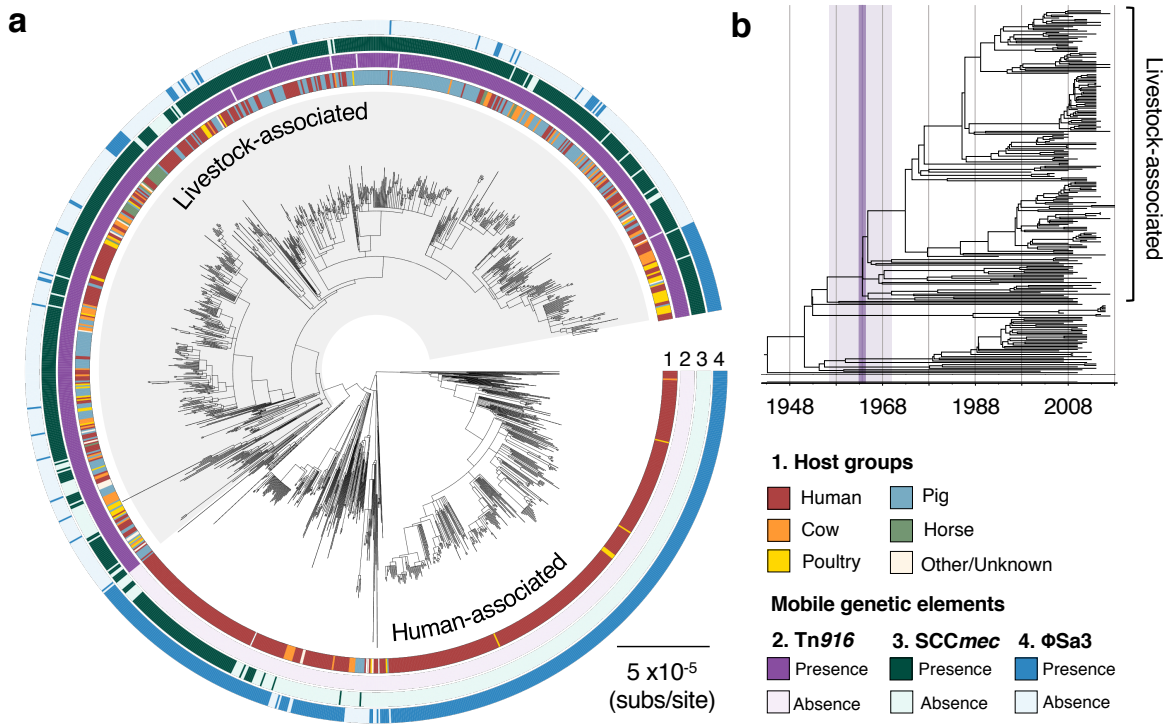


Figure 1. The transition to livestock-association in the 1960s was accompanied by changes in the frequencies of three mobile genetic elements. (a) A maximum likelihood phylogeny of 1,180 isolates of CC398, rooted using an outgroup from ST291. Grey shading indicates the livestock-associated clade. Outer 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 IEC. (b) A dated phylogeny of a sample of 250 CC398 isolates, showing that livestock-associated CC398 originated around 1964 (95% HPD: 1957-1970) (see Figures S3-S4 for details).

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 (Table S3, Figure 2a, Table S4) [21,22]. Several lines of evidence indicate that the presence of the Tn*916* transposon in livestock-associated CC398 is the result of a single integration event, followed by vertical 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, 95% CI: 51-64 years). Nevertheless, its absence from 5/704 livestock-associated CC398 isolates in our collection suggests that it remains capable of excision. 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 S6). These five isolates are broadly distributed across the livestock-associated clade, and not associated with a particular host species or geographic location.

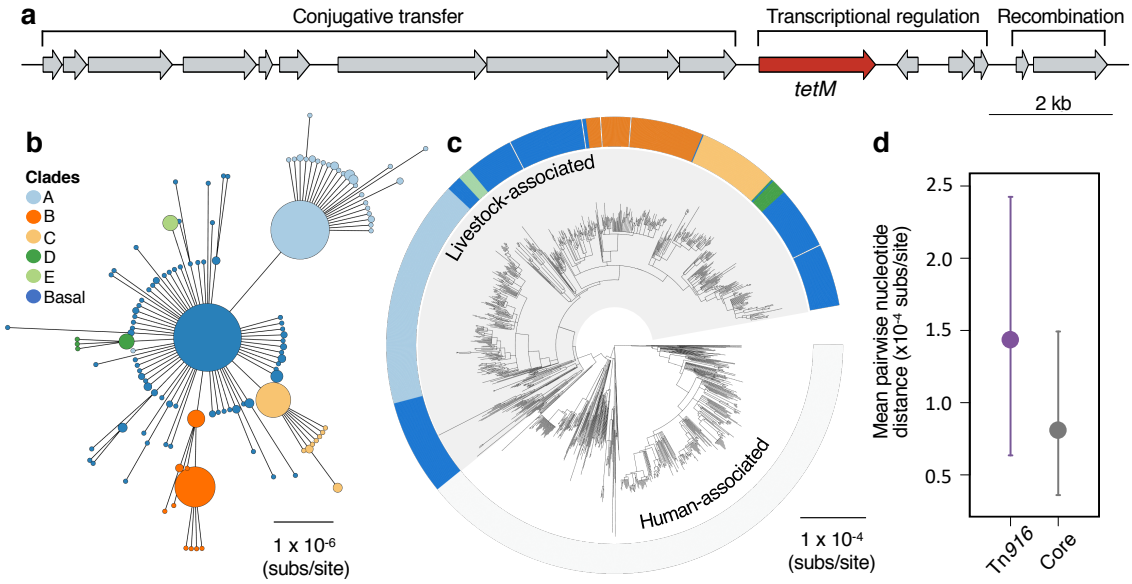


Figure 2. A Tn*916* transposon carrying *tetM* has been stably maintained by livestock-associated CC398 since its origin. (a) A gene map of the Tn*916* transposon in CC398 (based on reference genome 1_1439), with annotations based on previous studies [21,22]. (b) A minimum spanning tree of the element based on a concatenated alignment of all genes shown in (a). Points represent groups of identical elements, with point size correlated with group size on a log-scale, and colours representing well-supported clades (>70 bootstrap support in a maximum likelihood phylogeny) that include >10 isolates (smaller clades are incorporated into their basal clade). (c) These clades 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 have suggested that LA-MRSA CC398 emerged from human-associated MSSA [17]. 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 [23,24], Denmark [25], and New Zealand [26], makes the association between methicillin-resistance and livestock-association less clear (Figure 1a).

SCC*mec* elements in *S. aureus* are categorised into several types [27]. Consistent with previous studies of CC398 [e.g. 17], we observe both type V (76%) and type IV (21%) in CC398 (we were unable to confidently type the remaining 3%; Figures 3c, S8, Table S3). Most type V SCC*mec* elements belong to the subtype Vc previously described in livestock-associated CC398 [17,28,29]. This element includes two additional resistance genes: *tetK* (tetracycline resistance) and *czrC* (heavy metal resistance) (Figure 3a, Table S5). We identified a full-length version of this element in 335 genomes (including 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 were only observed in the livestock-associated clade, while shorter versions were also observed in the human-associated group. Shorter versions often lacked *tetK* (n=90) and *czrC* (n=117) genes (Figure S8). Type IV elements were 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) there is less of a difference: 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 S8). In contrast, diversity in the type IV elements in livestock-associated CC398 supports four independent acquisitions from outside of CC398 (Figure S9, Table S3). Similar to the type Vc element, once acquired, these elements tend to be maintained.

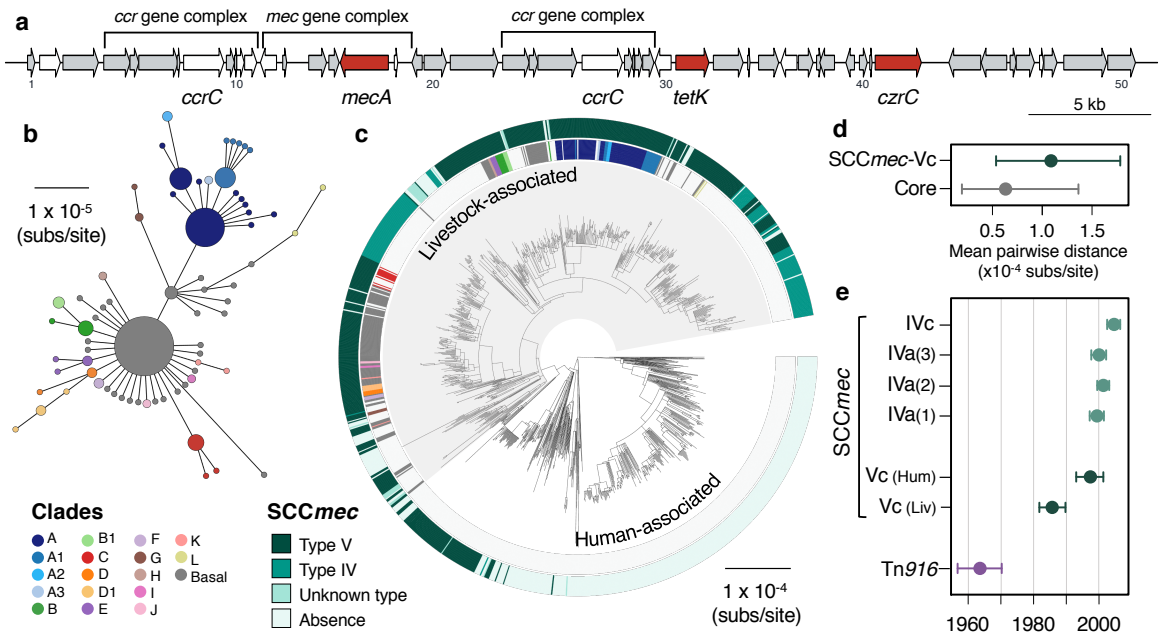


Figure 3. A type V SCC*mec* has been maintained by some lineages of livestock-associated CC398 since the 1980s, with occasional replacements with type IV SCC*mec* in the 1990s-2000s. (a) A gene map of the type Vc *SCCmec* element in CC398 (using the 1_1439 reference strain), with annotations from previous studies [28,29]. Genes in white were excluded from analyses of diversity within the element. (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 two external rings. (d) Mean pairwise nucleotide distance between isolates carrying the 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) Dates of acquisition of different SCC*mec* types and Tn*916* into CC398 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 [26]. 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 S8). 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 S8).

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 (Figures 3e and S10). These reconstructions consistently estimated that the type Vc SCC*mec* had been acquired by livestock-associated CC398 by approximately 1986 (95% HPD: 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 into human-associated CC398. In contrast, type IV elements show evidence of several more recent acquisitions in relatively quick succession between 1997 and 2004 (Figures 3e and S10).

Together, our analyses are consistent with LA-MRSA emerging from human-associated MSSA and acquiring the type Vc SSC*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 SSC*mec*.

3. Loss of a ϕSa3 prophage carrying a human IEC*, 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 IECs that include a variable set of functional genes that encode human-specific virulence factors, including *sak, chp*, *scn*, *sea* and *sep* [30–35]. They are temperate prophages that integrate into the *hlb* gene of *S. aureus*. While ϕSa3 prophages are present in 88% of our human-associated CC398 isolates, 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, Tables S3), 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) [30,34].

While ϕSa3 prophages are rare in livestock-associated CC398 (68/704 isolates), diversity within these elements indicate at least 15 acquisitions (but likely more) into livestock-associated CC398. The majority of these elements (69%) do not share a recent common ancestor with those in human-associated CC398. ϕSa3 prophages tend to have recent origins in livestock-associated CC398, with a notable exception being a previously described poultry-associated

sub-clade (n=51) [17,36–38]. our collection, 39% of the isolates in this clade are from poultry (compared to 4% poultry across the rest of the livestock-associated clade). Low nucleotide diversity within the type B ϕSa3 that is consistently found in isolates within this clade, suggests that it has been maintained since its acquisition approximately 21 years ago (95% CI: 1997-2001; Figure S11).

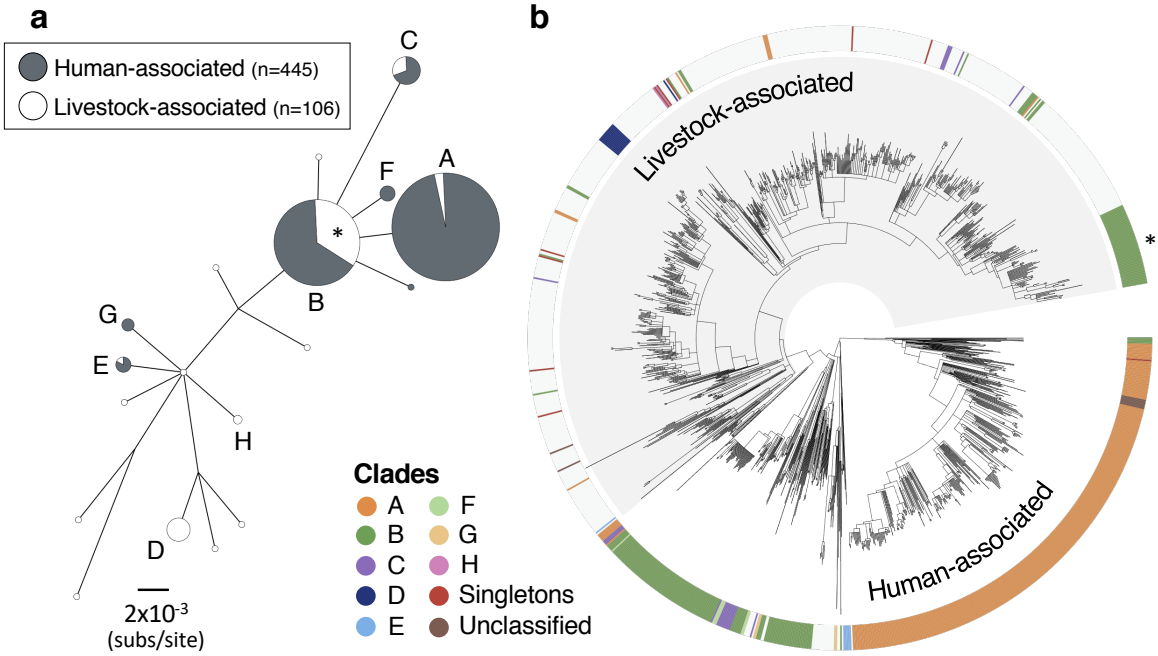


Figure 4. ϕSa3 prophages have been lost and reacquired multiple times in both human-associated and livestock-associated CC398. (a) A maximum likelihood phylogeny based on 12 genes shared across the ϕSa3 prophages in our collection, with both low-support nodes (<70% bootstrap support) and branches <0.0018 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 is correlated with the number of isolates 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 *.

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 isolates from livestock-associated CC398 carry one or more resistance genes for antibiotic classes other than tetracyclines and β-lactams (Figures S12 and S13, Table S7). 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 groups that include isolates from both human and livestock or companion animal hosts (Figure S14). Comparisons across these groups revealed that isolates from humans are consistently more likely to carry an ϕSa3 prophage (McNemar's Chi-squared test: $p = 0.001$; Figure 5).

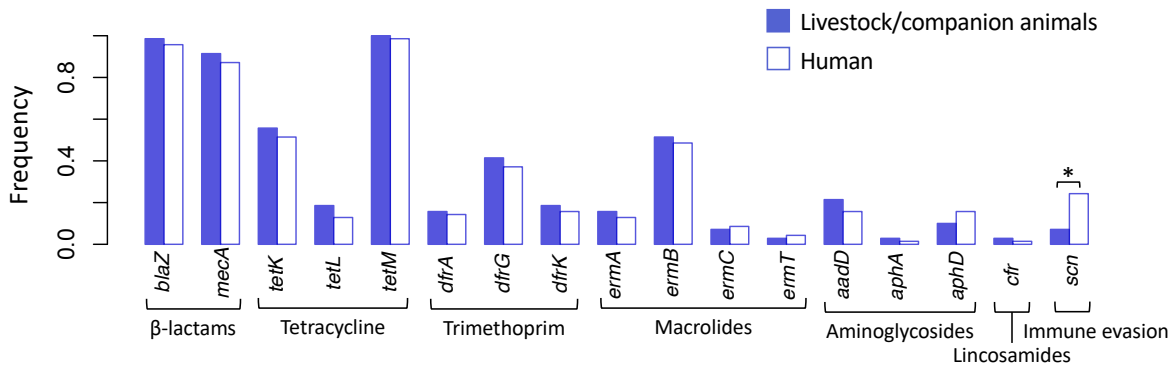


Figure 5. Spillover of livestock-associated CC398 into humans is associated with acquisition of human immune-evasion genes, but not a loss of antibiotic resistance genes. 70 phylogenetically independent clades that include isolates from both humans and other species were identified within the livestock-associated 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 ($p < 0.01$).

ϕSa3 prophages are also less common in pigs than other non-human species (only 3% of pig isolates carry an ϕSa3 prophage). 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 were 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 antibiotic resistance genes (Figure 5). No resistance gene was significantly more or less common in humans than in other species, nor was there a consistent shift in the overall number of resistance genes.

Discussion

We have characterised the 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 IEC are rare in livestock-associated CC398 and common in human-associated CC398, there have been frequent gains and losses in both livestock- and human-associated CC398. These contrasting dynamics may reflect variation in the selective benefits, selective costs, and availability of these MGEs.

The Tn*916* transposon carrying *tetM* integrated at the origin of CC398 in livestock in 1964 (95% CI: 1957-1970). The stable integration of this transposon could have been of considerable benefit to a livestock-associated lineage since tetracyclines are the most commonly used antimicrobial class in livestock farming [9,16]. In addition, previous studies have suggested that Tn*916* is associated with a low selective burden in the absence of treatment, which could increase its stability [21]. While Tn*916* is found across several bacterial genera [21,39], the element is relatively rare in *S. aureus*; while it has been observed in other CCs (5, 40 and 35), there is no evidence of its horizontal transmission within *S. aureus* [22]. This combination of high selective benefit, low selective cost and rarity could explain the remarkable stability of Tn*916* in livestock-associated CC398, and may partly 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 [40]. *czrC* is associated with heavy metal resistance [41], 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 [42]. Additionally, *mecA* is likely to be beneficial because of the common use of beta-lactams in livestock farming, including third generation cephalosporins [43,44]. 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 [45]), but has not been reported in other *S. aureus* CCs.

Losses of the type Vc SCC*mec* on internal branches within livestock-associated CC398 are 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 dynamics of the type Vc SCC*mec* are broadly consistent with a high selective benefit (particularly in pigs), a high cost, and the rarity of this element relative to other SCC*mec*.

While the loss of the ϕSa3 prophage 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 8], is consistent with a strong benefit in the human host. On the other hand, as we find that these elements are frequently lost, and are generally absent in other host species, these elements likely carry a selective burden outside of the human host. The diversity of ϕSa3 prophages present in CC398 suggests they form a large pool of elements [30] that are readily acquired by *S. aureus* in the human host.

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 [17,36–38]. This clade was first described as a hybrid LA-MRSA CC9/CC398 lineage [17], and was subsequently investigated as a lineage associated with human disease [38]. 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 these elements might aid immune evasion in species other than humans [46]). 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 are 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.

The stability of Tn*916* and SCC*mec* in livestock-associated CC398 suggests that these elements are capable of persisting across different livestock species and countries, and therefore of maintenance across variable levels and types of antibiotic exposure. While the SCC*mec* is less stably maintained than the Tn*916* transposon, 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 gradual reductions in antibiotic consumption and the forthcoming EU ban on medical zinc supplementation in pig feed [9,47–49], may not lead to the loss of these elements. However, further work is needed to fully resolve the factors that underlie the acquisition and maintenance of MGEs carrying 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 (Table S1). We additionally sequenced 5 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 [53].

Genome assembly

Sequence data from all isolates was used to generate *de novo* assemblies with *Spades* v.3.12.0 [51]. Adapters and low quality reads were removed with *Cutadapt* v1.16 [52] and *Sickle* v1.33 [53], and screened for contamination using *FastQ Screen* v0.12.0. Optimal k-mers were identified based on average read lengths for each genome. All assemblies were evaluated using QUAST v.5.0.1 [54] and reads were mapped back to *de novo* assemblies to investigate polymorphism (indicative of mixed cultures) using *Bowtie2* v1.2.2 [55]. 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). Genomes mischaracterised as CC398 were identified via two approaches and excluded from further analysis. First, sequence types (STs) were identified with *MLST-check* [56] and grouped into CCs using the *eBURST* algorithm with a single locus variant [57]. 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 [58].

Reference-mapped assemblies were generated using *Bowtie2* using the reference genome S0385 (GenBank accession no. AM990992). For reference genomes, artificial FASTQ files were

generated using *ArtificialFastqGenerator* [59]. 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.

Recombination was identified in the reference-mapped alignment using both *Gubbins* v2.3.1 [60] and *ClonalFrame* [61]. All recombinant sites identified were masked 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 [17].

Genome annotation and identification of homologous genes

De novo assemblies were annotated with *Prokka* (v2.8.2) [62] and orthologous genes identified with *Roary* [63] using recommended parameter values. A core gene alignment was created with *Roary* and recombinant sites identified and masked using *Gubbins*. Antibiotic resistance genes were identified using the Pathogenwatch AMR prediction module (Wellcome Sanger Institute), which uses BLASTn [64] 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

Phylogenetic reconstruction was carried out for the reference-mapped alignment with *RAxML* v8.2.4 using the GTR+Γ model and 1,000 bootstraps [65]. Sites where >0.1% of genomes showed evidence of recombination or had missing data were excluded from the analysis. Dated phylogenies were constructed 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 [66]. Separate substitution models and molecular clocks were fit to 1st/2nd and 3rd codon positions to reflect differences in selective constraint. Phylogenies were constructed for three random subsamples of 250 isolates (200 from livestock-associated CC398 and 50 from human-associated CC398). Subsamples were non-overlapping with the exception of 30 genomes representing the most divergent lineages within the livestock-associated clade, to ensure a consistent description the origin of this clade. Additional phylogenies were constructed from subsamples that included only isolates from the LA-clade to establish consistent rate estimates were returned over different evolutionary depths. Temporal signal was investigated in each data set using a regression of root-to-tip distance against sampling date, and a permutation test of dates over tips (with clustering used to correct for confounding of temporal and genetic structures) [67,68]. Trees were visualized and annotated using ITOL [69].

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. The association between these genes and MGEs was investigated 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.

The identity of the Tn*916* transposon was confirmed 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 SCC*mec*Finder database in our *de novo* assemblies (Table S7) [27]. ϕSa3 prophages were initially identified and categorised through identification of functional genes associated with the human IEC. 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 [70], 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* [58], and constructed maximum likelihood trees using RaxML and minimum-spanning trees using *GrapeTree* [71] 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.

Phylogenetically independent groups contrast

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 groups 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 within that group.

Authors' contributions

MM collected and assembled all the data used in this study. XB and RW sampled and sequenced the five new isolates from UK farms. LAW and MAH conceived the project. MM and GGRM designed and performed the phylogenetic and comparative genomic analyses. GGRM, MM and LAW wrote the manuscript, with input from all other authors.

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

1. Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? *Heredity* 106:1, 1–10. doi:10.1038/hdy.2010.24

2. Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* 3:9. 722–732. doi:10.1038/nrmicro1235

3. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The Ecology and Evolution of Pangenomes. *Current Biology*. 29: R1094–R1103. doi:10.1016/J.CUB.2019.08.012

4. Turner NA, Sharma-Kuinkel BK, Maskarinec SA, Eichenberger EM, Shah PP, Carugati M, et al. 2019. Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nature Reviews Microbiology* 17:4, 203–218. doi:10.1038/s41579-018-0147-4

5. Matuszewska M, Murray GGR, Harrison EM, Holmes MA, Weinert LA. 2020. The Evolutionary Genomics of Host Specificity in *Staphylococcus aureus*. *Trends in Microbiology*. 28:6, 465-477. doi:10.1016/j.tim.2019.12.007

6. Haag AF, Fitzgerald JR, Penadés JR. 2019. *Staphylococcus aureus* in Animals. *Microbiology Spectrum*. 7. doi:10.1128/MICROBIOLSPEC.GPP3-0060-2019

7. Jamrozy D, Coll F, Mather AE, Harris SR, Harrison EM, MacGowan A, et al. 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* 18:1, 1–12. doi:10.1186/S12864-017-4065-Z

8. Sieber RN, Larsen AR, Urth TR, Iversen S, Møller CH, Skov RL, et al. 2019. Genome investigations show host adaptation and transmission of LA-MRSA CC398 from pigs into Danish healthcare institutions. *Scientific Reports* 9:1, 1–10. doi:10.1038/s41598-019-55086-x
9. DANMAP reports 2018. Available: https://www.danmap.org/Reports/2018
10. Lekkerkerk WSN, van Wamel WJB, Snijders S v., Willems RJ, van Duijkeren E, Broens EM, et al. 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. doi:10.1128/JCM.02702-14
11. Islam MZ, Espinosa-Gongora C, Damborg P, Sieber RN, Munk R, Husted L, et al. 2017. Horses in Denmark Are a Reservoir of Diverse Clones of Methicillin-Resistant and -Susceptible *Staphylococcus aureus*. *Frontiers in Microbiology*. 543. doi:10.3389/FMICB.2017.00543
12. Anjum MF, Marco-Jimenez F, Duncan D, Marín C, Smith RP, Evans SJ. 2019. Livestock-Associated Methicillin-Resistant *Staphylococcus aureus* From Animals and Animal Products in the UK. *Frontiers in Microbiology*.10. doi:10.3389/FMICB.2019.02136
13. Larsen J, Petersen A, Larsen AR, Sieber RN, Stegger M, Koch A, et al. Emergence of Livestock-Associated Methicillin-Resistant *Staphylococcus aureus* Bloodstream Infections in Denmark. 2017. *Clinical Infectious Diseases*. 65:1072. doi:10.1093/CID/CIX504
14. van Alen S, Ballhausen B, Peters G, Friedrich AW, Mellmann A, Köck R, et al. 2017. In the centre of an epidemic: Fifteen years of LA-MRSA CC398 at the University Hospital Münster. *Veterinary Microbiology*. 200:19–24. doi:10.1016/J.VETMIC.2016.01.021
15. Cantas L, Suer K. Review: The Important Bacterial Zoonoses in “One Health” Concept. 2014. *Frontiers in Public Health*. 0: 144. doi:10.3389/FPUBH.2014.00144
16. WHO. Global Action Plan on Antimicrobial Resistance. 2015.
17. Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, et al. 2012. *Staphylococcus aureu*s CC398: Host adaptation and emergence of methicillin resistance in livestock. *mBio*. 3: 1–6. doi:10.1128/MBIO.00305-11
18. Ward MJ, Gibbons CL, McAdam PR, van Bunnik BAD, Girvan EK, Edwards GF, et al. 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. doi:10.1128/AEM.01777-14
19. Silva AG da, Baines SL, Carter GP, Heffernan H, French NP, Ren X, et al. 2017. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant *Staphylococcus aureus*. *Microbial Genomics*. 3. doi:10.1099/MGEN.0.000105

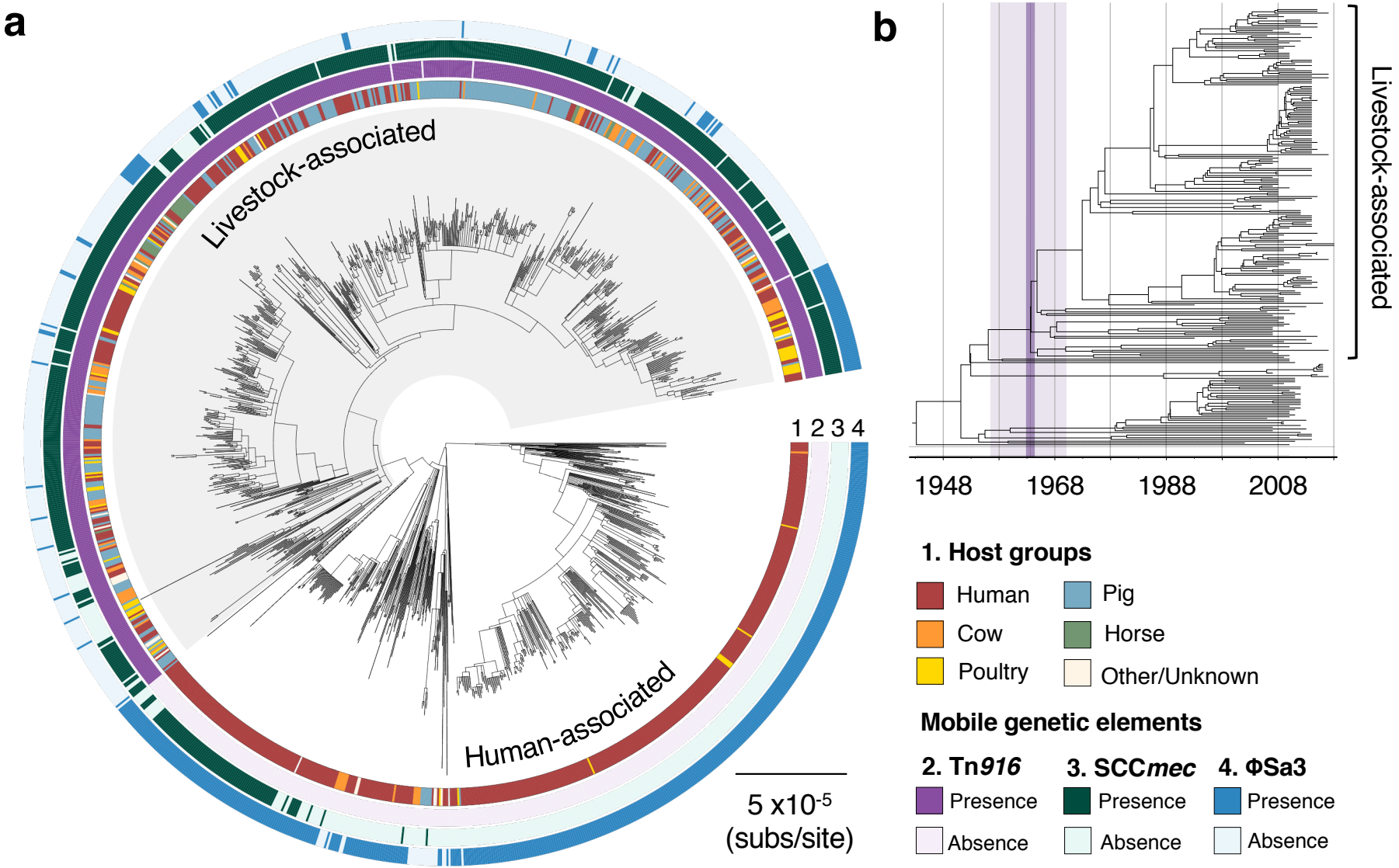
20. Hsu L-Y, Harris SR, Chlebowicz MA, Lindsay JA, Koh T-H, Krishnan P, et al. 2015. Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. *Genome Biology*. 16:1. 1–13. doi:10.1186/S13059-015-0643-Z
21. Roberts AP, Mullany P. 2009. A modular master on the move: the Tn*916* family of mobile genetic elements. *Trends in Microbiology*. 17: 251–258. doi:10.1016/J.TIM.2009.03.002
22. de Vries LE, Christensen H, Skov RL, Aarestrup FM, 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. doi:10.1093/JAC/DKP214
23. He L, Zheng H-X, Wang Y, Le KY, Liu Q, Shang J, et al. 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. doi:10.1186/S13073-018-0514-9
24. Zou G, Matuszewska M, Jia M, Zhou J, Ba X, Duan J, et al. 2021. A survey of Chinese pig farms and human healthcare isolates reveals separate human and animal MRSA populations. *bioRxiv*. doi:10.1101/2021.08.02.454852
25. Møller JK, Larsen AR, Østergaard C, Møller CH, Kristensen MA, Larsen J. 2019. International travel as source of a hospital outbreak with an unusual methicillin-resistant *Staphylococcus aureus* clonal complex 398, Denmark, 2016. *Eurosurveillance*. 24: 1800680. doi:10.2807/1560-7917.ES.2019.24.42.1800680
26. Silva AG da, Baines SL, Carter GP, Heffernan H, French NP, Ren X, et al. 2017. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant *Staphylococcus aureus*. *Microbial Genomics*. 3. doi:10.1099/MGEN.0.000105
27. Hanssen A-M, Ericson Sollid JU. 2006. SCC*mec* in staphylococci: genes on the move. *FEMS Immunology* & Medical Microbiology. 46: 8–20. doi:10.1111/J.1574-695X.2005.00009.X
28. Li S, Skov RL, Han X, Larsen AR, Larsen J, Sørum M, et al. 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. doi:10.1128/AAC.01475-10
29. Vandendriessche S, Vanderhaeghen W, Larsen J, de Mendonça R, Hallin M, Butaye P, et al. 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. doi:10.1093/JAC/DKT366
30. 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

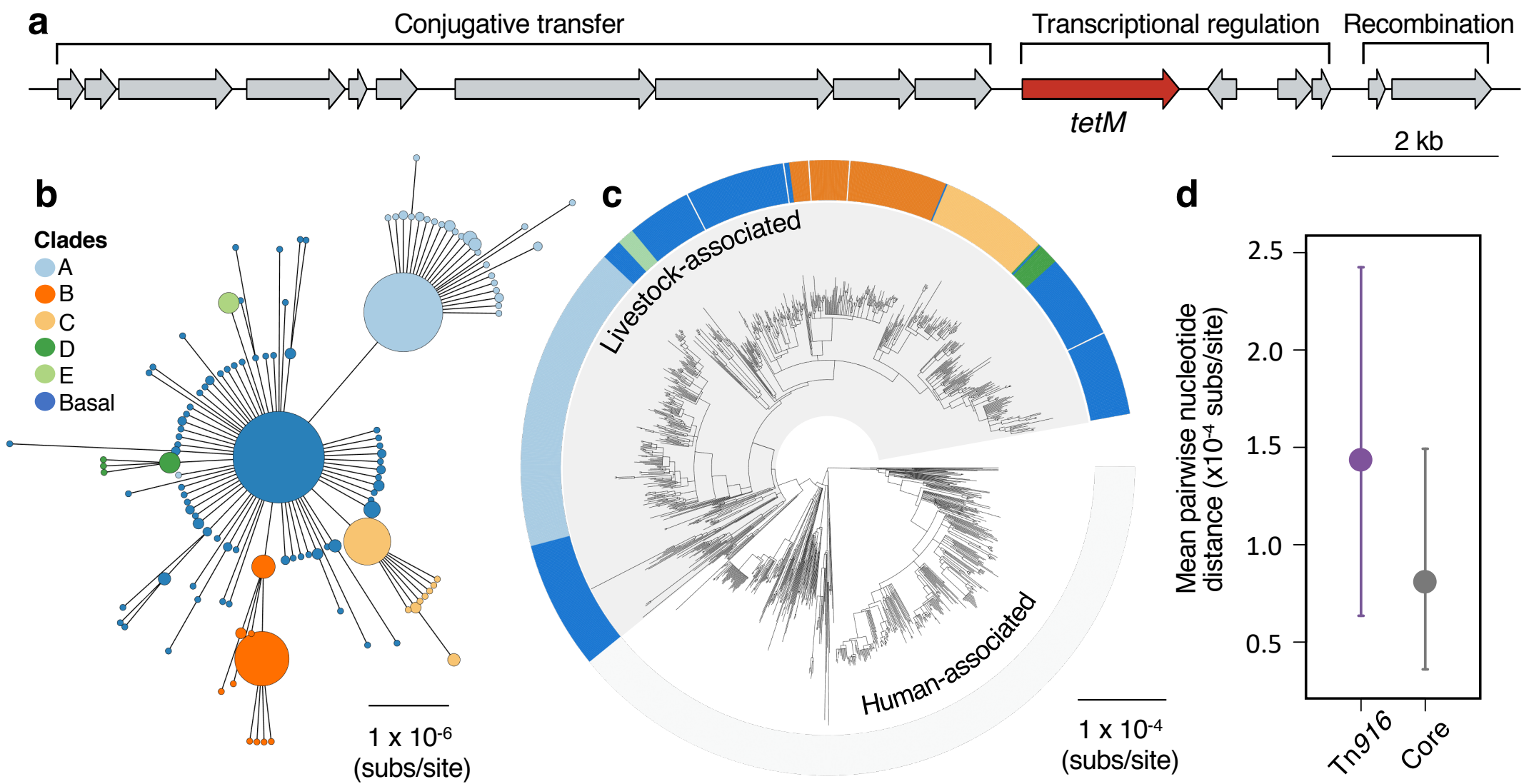
Staphylococcus aureus isolates from 2000 to 2015. *Journal of Clinical Microbiology*. 56. doi:10.1128/JCM.00140-18
31. Gladysheva IP, Turner RB, Sazonova IY, Liu L, Reed GL. 2003. Coevolutionary patterns in plasminogen activation. *Proceedings of the National Academy of Sciences*.100: 9168–9172. doi:10.1073/PNAS.1631716100
32. Postma B, Poppelier MJ, Galen JC van, Prossnitz ER, Strijp JAG van, Haas CJC de, et al. 2004. Chemotaxis Inhibitory Protein of *Staphylococcus aureus* Binds Specifically to the C5a and Formylated Peptide Receptor. *The Journal of Immunology*.172: 6994–7001. doi:10.4049/JIMMUNOL.172.11.6994
33. Rooijakkers SHM, Ruyken M, Roos A, Daha MR, Presanis JS, Sim RB, et al. 2005. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nature Immunology*. 6:9. 6: 920–927. doi:10.1038/ni1235
34. van Wamel WJB, Rooijakkers SHM, Ruyken M, van Kessel KPM, van Strijp JAG. 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. doi:10.1128/JB.188.4.1310-1315.2006
35. Thammavongsa V, Kim HK, Missiakas D, Schneewind O. 2015. Staphylococcal manipulation of host immune responses. *Nature Reviews Microbiology* 13:9. 2015;13: 529–543. doi:10.1038/nrmicro3521
36. Pérez-Moreno MO, Centelles-Serrano MJ, Nogales-López J, Domenech-Spanedda MF, 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. doi:10.1016/J.EIMC.2016.07.008
37. Tang Y, Larsen J, Kjeldgaard J, Andersen PS, Skov R, Ingmer H. 2017. Methicillin-resistant and -susceptible *Staphylococcus aureus* from retail meat in Denmark. *International Journal of Food Microbiology*. 249: 72–76. doi:10.1016/J.IJFOODMICRO.2017.03.001
38. Larsen J, Stegger M, Andersen PS, Petersen A, Larsen AR, Westh H, et al. 2016. Evidence for Human Adaptation and Foodborne Transmission of Livestock-Associated Methicillin-Resistant *Staphylococcus aureus. Clinical Infectious Diseases*. 63: 1349–1352. doi:10.1093/CID/CIW532
39. Clewell DB, Flannagan SE, Jaworski DD, Clewell DB. 1995. Unconstrained bacterial promiscuity: the Tn916–Tn1545 family of conjugative transposons. *Trends in Microbiology*. 3: 229–236. doi:10.1016/S0966-842X(00)88930-1
40. Larsen J, Clasen J, Hansen JE, Paulander W, Petersen A, Larsen AR, et al. 2016. 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

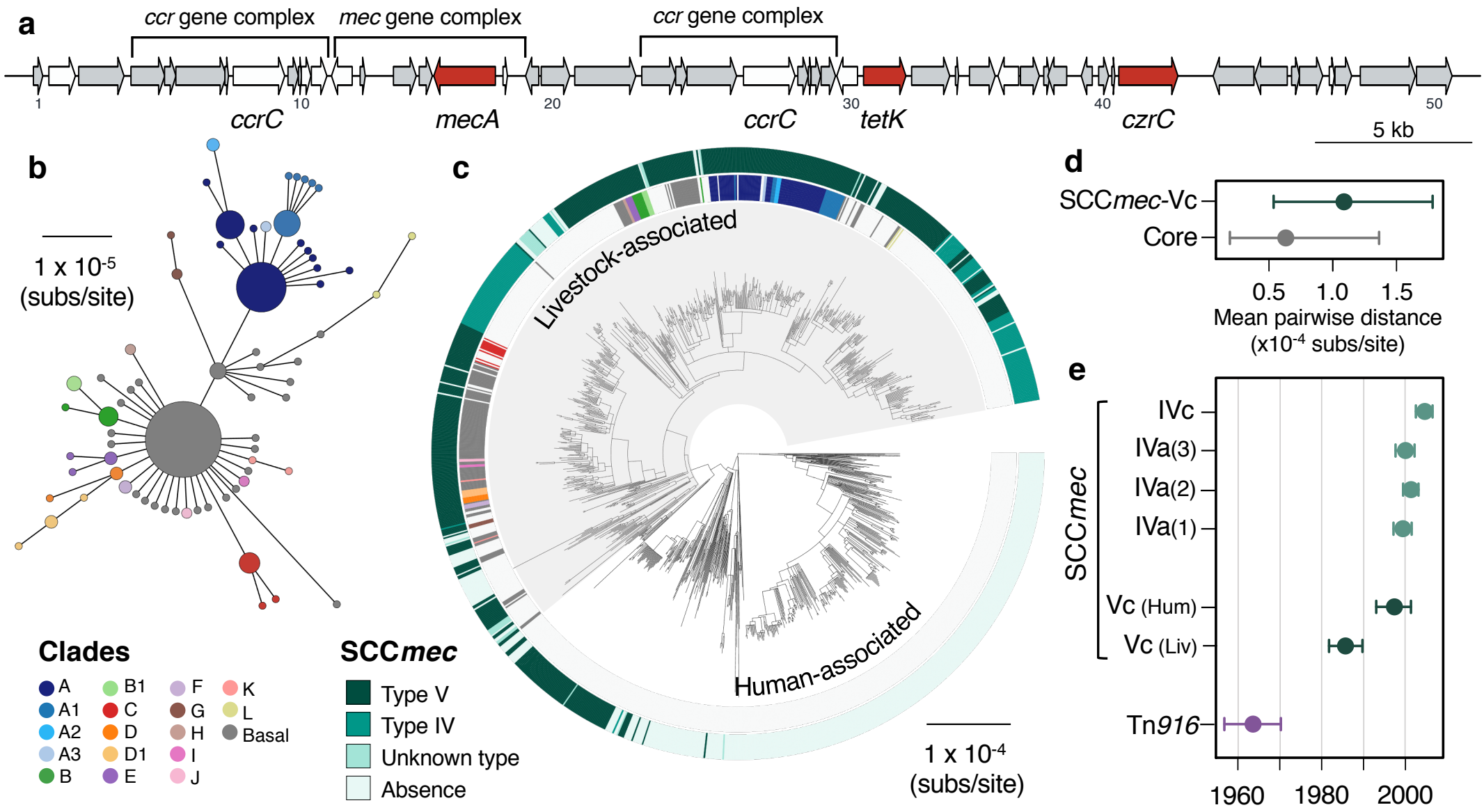
concentrations of tetracycline. *Antimicrobial Agents and Chemotherapy*. 60: 4401–4403. doi:10.1128/AAC.00426-16
41. Cavaco LM, Hasman H, Stegger M, Andersen PS, Skov R, Fluit AC, et al. 2010. Cloning and occurrence of czrC, a gene conferring cadmium and zinc resistance in methicillin-resistant *Staphylococcus aureus* CC398 isolates. Antimicrobial Agents and Chemotherapy. 54: 3605–3608. doi:10.1128/AAC.00058-10
42. Nielsen CL, Kongsted H, Sørensen JT, Krogh MA. 2021. Antibiotic and medical zinc oxide usage in Danish conventional and welfare-label pig herds in 2016–2018. *Preventive Veterinary Medicine*. 189: 105283. doi:10.1016/J.PREVETMED.2021.105283
43. Sjölund M, Postma M, Collineau L, Lösken S, Backhans A, Belloc C, et al. 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. doi:10.1016/J.PREVETMED.2016.06.003
44. Lekagul A, Tangcharoensathien V, Yeung S. 2019. Patterns of antibiotic use in global pig production: A systematic review. *Veterinary and Animal Science*. 7: 100058. doi:10.1016/J.VAS.2019.100058
45. Hoefer A, Boyen F, Beierschmitt A, Moodley A, Roberts MC, Butaye P. 2021. Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus* from Vervet Monkeys (*Chlorocebus sabaeus*) in Saint Kitts. *Antibiotics*.10. doi:10.3390/ANTIBIOTICS10030290
46. Jung P, Abdelbary MMH, Kraushaar B, Fetsch A, Geisel J, Herrmann M, et al. 2017. Impact of bacteriophage Saint3 carriage on the immune evasion capacity and hemolytic potential of *Staphylococcus aureus* CC398. *Veterinary Microbiology*. 200: 46–51. doi:10.1016/J.VETMIC.2016.02.015
47. Zinc oxide, European Medicines Agency. Available: https://www.ema.europa.eu/en/medicines/veterinary/referrals/zinc-oxide
48. European Surveillance of Veterinary Antimicrobial Consumption (ESVAC), European Medicines Agency. Available: https://www.ema.europa.eu
49. UK One Health Report: antibiotic use and antibiotic resistance in animals and humans - GOV.UK. Available: https://www.gov.uk/government/publications/uk-one-health-report-antibiotic-use-and-antibiotic-resistance-in-animals-and-humans
50. Harrison EM, Paterson GK, Holden MTG, Larsen J, Stegger M, Larsen AR, et al. 2013. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. *EMBO Molecular Medicine*. 5: 509–515. doi:10.1002/EMMM.201202413
51. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. 19: 455. doi:10.1089/CMB.2012.0021

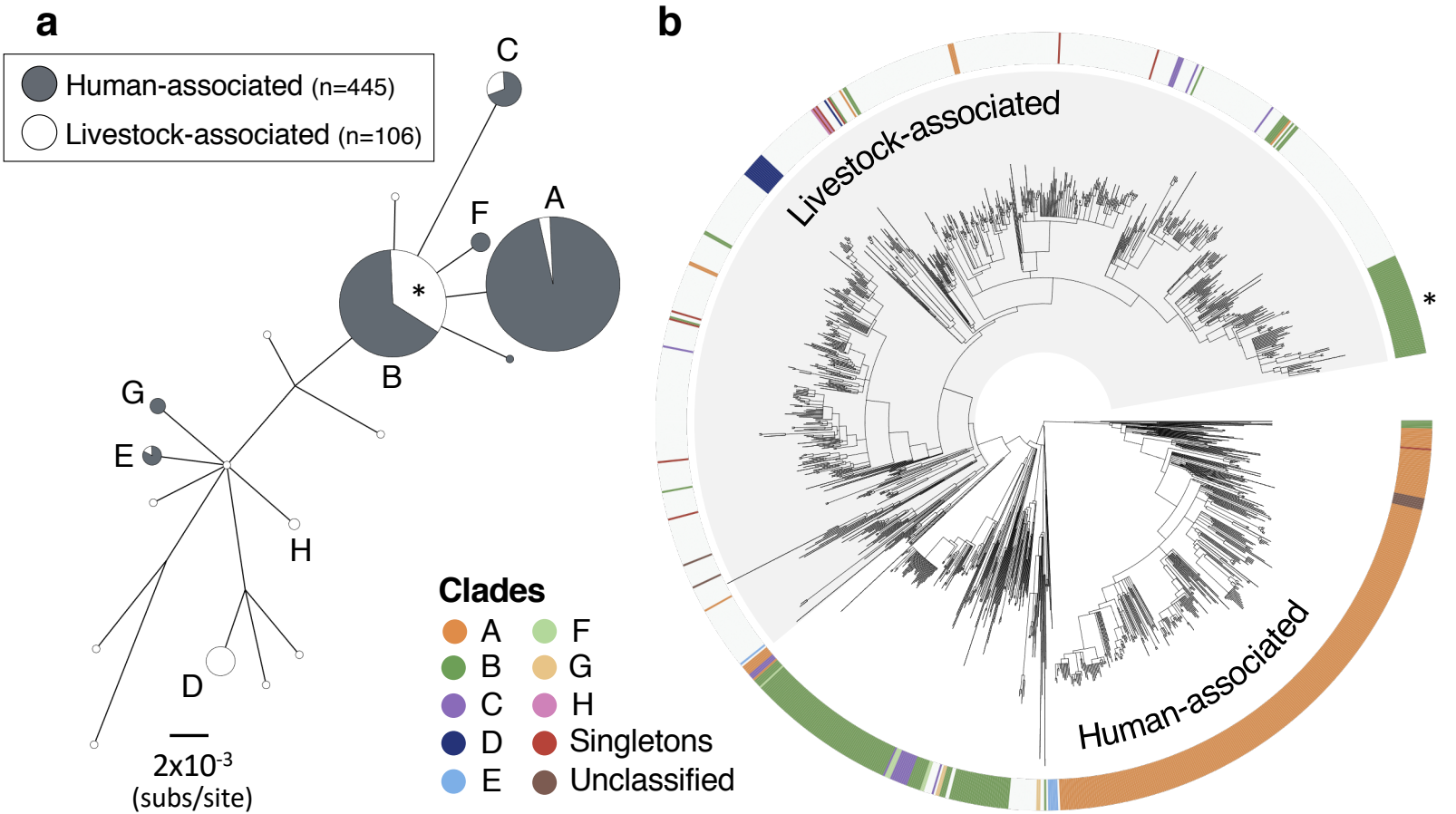
52. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*.17: 10–12. doi:10.14806/EJ.17.1.200
53. Joshi N, Fass J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. [software]
54. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29: 1072–1075. doi:10.1093/BIOINFORMATICS/BTT086
55. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9:4, 357–359. doi:10.1038/nmeth.1923
56. Page AJ, Silva N de, Hunt M, Quail MA, Parkhill J, Harris SR, et al. 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microbial Genomics*. 2: e000083. doi:10.1099/MGEN.0.000083
57. Francisco AP, Bugalho M, Ramirez M, Carriço JA. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 10:1,1–15. doi:10.1186/1471-2105-10-152
58. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*.20: 289–290. doi:10.1093/bioinformatics/btg412
59. Frampton M, Houlston R. 2012. Generation of Artificial FASTQ Files to Evaluate the Performance of Next-Generation Sequencing Pipelines. *PLOS ONE*. 7: e49110. doi:10.1371/JOURNAL.PONE.0049110
60. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research*. 43: e15–e15. doi:10.1093/NAR/GKU1196
61. Didelot X, Wilson DJ. 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Computational Biology*.11. doi:10.1371/JOURNAL.PCBI.1004041
62. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 30: 2068–2069. doi:10.1093/bioinformatics/btu153
63. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 31: 3691–3693. doi:10.1093/BIOINFORMATICS/BTV421
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215: 403–410. doi:10.1016/S0022-2836(05)80360-2
65. Stamatakis A. 2014. RAxML v.8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30: 1312–1313. doi:10.1093/BIOINFORMATICS/BTU033
66. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*. 29: 1969–1973. doi:10.1093/MOLBEV/MSS075

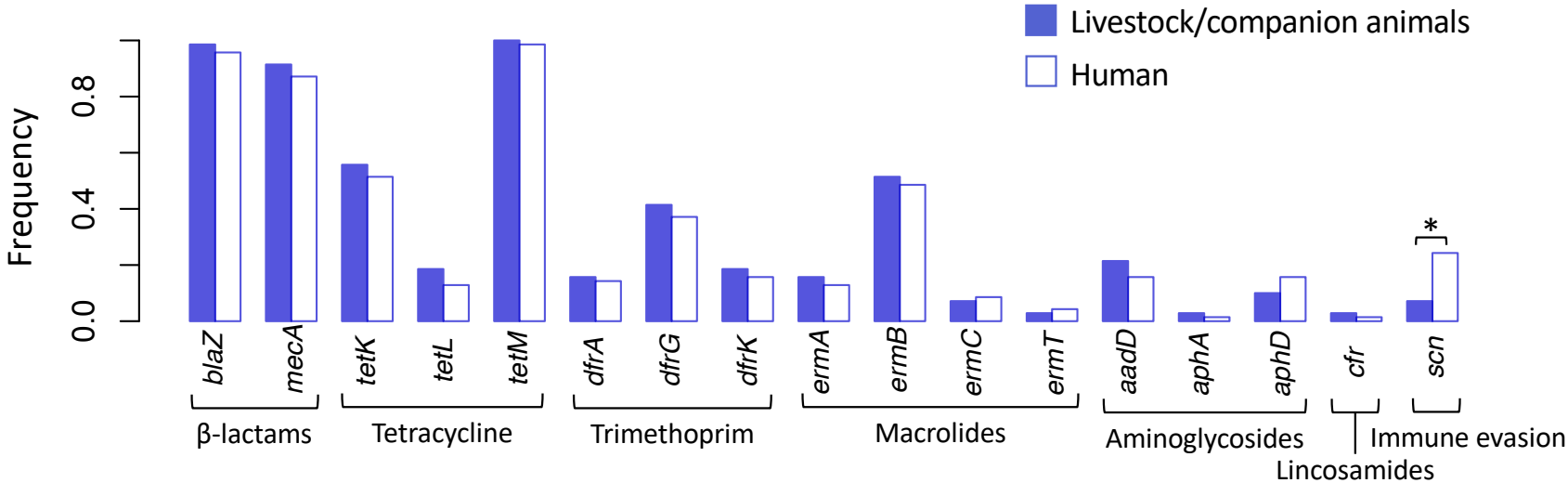
67. Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, et al. 2016. The effect of genetic structure on molecular dating and tests for temporal signal. *Methods in Ecology and Evolution*. 7: 80–89. doi:10.1111/2041-210X.12466
68. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*.2. doi:10.1093/VE/VEW007
69. 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. doi:10.1093/NAR/GKAB301
70. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 27: 135–145. doi:10.1002/PRO.3290
71. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Research*. 28: 1395. doi:10.1101/GR.232397.117


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
Presence
Presence
Absence
Absence
Absence


a
Conjugative transfer
Transcriptional regulation
Recombination
tetM
2 kb
b
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)
Clades
A
A1
A2
A3
B
B1
C
D
D1
E
F
G
H
I
J
K
L
Basal
c
Livestock-associated
Human-associated
1 x 10-4
(subs/site)
SCCmec
Type V
Type IV
Unknown type
Absence
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 (Hum)
Vc (Liv)
Tn916
1960
1980
2000


a
Human-associated (n=445)
Livestock-associated (n=106)
C
A
F
*
B
G
E
H
D
2x10-3
(subs/site)
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