



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* CC398**

**Citation for published version:**

Ward, MJ, Gibbons, CL, McAdam, PR, van Bunnik, BAD, Girvan, EK, Edwards, GF, Fitzgerald, R & Woolhouse, MEJ 2014, 'Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* CC398' *Applied and Environmental Microbiology*, vol. 80, no. 23. DOI: 10.1128/AEM.01777-14

**Digital Object Identifier (DOI):**

[10.1128/AEM.01777-14](https://doi.org/10.1128/AEM.01777-14)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Applied and Environmental Microbiology*

**Publisher Rights Statement:**

Copyright © 2014 Ward et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Time-Scaled Evolutionary Analysis of the Transmission and Antibiotic Resistance Dynamics of *Staphylococcus aureus* Clonal Complex 398

M. J. Ward,<sup>a</sup> C. L. Gibbons,<sup>a</sup> P. R. McAdam,<sup>b</sup> B. A. D. van Bunnik,<sup>a</sup> E. K. Girvan,<sup>c</sup> G. F. Edwards,<sup>c</sup> J. R. Fitzgerald,<sup>b</sup> M. E. J. Woolhouse<sup>a</sup>

Centre for Immunity, Infection and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom<sup>a</sup>; The Roslin Institute and Edinburgh Infectious Diseases, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, United Kingdom<sup>b</sup>; Scottish MRSA Reference Laboratory, National Health Service Greater Glasgow and Clyde, Glasgow Royal Infirmary, Glasgow, United Kingdom<sup>c</sup>

***Staphylococcus aureus* clonal complex 398 (CC398) is associated with disease in humans and livestock, and its origins and transmission have generated considerable interest. We performed a time-scaled phylogenetic analysis of CC398, including sequenced isolates from the United Kingdom (Scotland), along with publicly available genomes. Using state-of-the-art methods for mapping traits onto phylogenies, we quantified transitions between host species to identify sink and source populations for CC398 and employed a novel approach to investigate the gain and loss of antibiotic resistance in CC398 over time. We identified distinct human- and livestock-associated CC398 clades and observed multiple transmissions of CC398 from livestock to humans and between countries, lending quantitative support to previous reports. Of note, we identified a subclade within the livestock-associated clade comprised of isolates from hospital environments and newborn babies, suggesting that livestock-associated CC398 is capable of onward transmission in hospitals. In addition, our analysis revealed significant differences in the dynamics of resistance to methicillin and tetracycline related to contrasting historical patterns of antibiotic usage between the livestock industry and human medicine. We also identified significant differences in patterns of gain and loss of different tetracycline resistance determinants, which we ascribe to epistatic interactions between the resistance genes and/or differences in the modes of inheritance of the resistance determinants.**

The bacterium *Staphylococcus aureus* can colonize and cause infection in humans, livestock, and other animals. Methicillin-resistant *S. aureus* (MRSA) presents a significant challenge to public health and veterinary medicine, and of particular concern is the emergence of strains that are resistant to multiple classes of antibiotics. MRSA can be acquired by humans in health care and community settings (1) and is an emerging cause of infection among individuals (and their close contacts) working with livestock (2).

*S. aureus* clonal complex 398 (CC398), as defined by multi-locus sequence typing (MLST) (3), includes MRSA and methicillin-sensitive *S. aureus* (MSSA) and has been reported in humans and livestock across the globe. While exposure to livestock has been identified as a risk factor for human CC398 infection (2), there have been human cases with no reported livestock contact (4), and the transmission of CC398 between humans has been inferred from epidemiological analyses (5, 6). In humans, invasive CC398 infection has been associated with a range of clinical outcomes, which may be life threatening (7). CC398 may also cause clinical infection in animals and has been associated with bovine mastitis (8).

Studies that have combined MLST sequence data from a number of sequence types (STs) have shown that *S. aureus* may jump between human and livestock hosts in both directions (9, 10). In an analysis of MLST sequences for almost 700 STs, ST398 clustered with STs that were wholly associated with human hosts (10), and a human ancestral host for ST398 was thus parsimoniously inferred. Microarray studies (6, 11) of CC398 have shown a distinction between livestock-associated infections and those associated with human-to-human transmission in terms of their distribution of mobile genetic elements (MGEs), which has been echoed by studies of the CC398 core genome (12). However, the evolutionary history of CC398 has not yet been studied with quan-

titative phylogeny-based analyses that explicitly incorporate information about the date, host, and location of sampling.

In contrast to mainland Europe, few publications have described CC398 colonization or infection of humans or animals in the United Kingdom. However, cases of human CC398 infection have been identified by the Scottish MRSA Reference Laboratory (SMRSARL), and CC398 has been reported in two horses in southeast England (13), in bulk tank milk from dairy cattle (14), and in United Kingdom poultry (15). Despite the close association between CC398 and pig farming, in 2008 it was reported that no MRSA isolate, either ST398 or non-ST398, was found on sampled holdings with breeding pigs in the United Kingdom (16). Strains from humans and animals in the United Kingdom have not been included in published genomic analyses of global CC398.

In this study, we performed a time-scaled global molecular epidemiological analysis of whole-genome *S. aureus* CC398 sequences from humans and livestock, and we discuss the phylogenetic placement of sequenced strains from Scotland. We used quantitative phylogenetic methods to identify sink and source populations for CC398 and formally investigated the evolutionary histories of antibiotic resistance in a phylogenetic framework.

Received 29 May 2014 Accepted 11 September 2014

Published ahead of print 19 September 2014

Editor: C. A. Elkins

Address correspondence to M. J. Ward, melissa.ward@ed.ac.uk.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.01777-14>.

Copyright © 2014 Ward et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

doi:10.1128/AEM.01777-14

## MATERIALS AND METHODS

**Identification of CC398 isolates from the Scottish *S. aureus* database.** Thirty-six CC398 isolates were identified from the SMRSARL database, resulting from surveillance between February 1997 and December 2011. SMRSARL receives *S. aureus* isolates (both MSSA and MRSA) from diagnostic laboratories across Scotland. In addition to isolates from humans, there was also an environmental (cleaning) isolate from a hospital (17) and an isolate from a veterinary laboratory. Isolates were identified as CC398 by searching for CC398-associated MLSTs and *spa* types or, where MLST and *spa* typing had not been carried out, pulsed-field gel electrophoresis (PFGE) profiles and PCR-ribotyping (PCR-R) profiles matching strains already identified as CC398. In the subsequent genomic analysis, membership in CC398 was confirmed by assessing the phylogenetic clustering with sequences from STs belonging to CC398.

The earliest CC398 record in the SMRSARL database was from 2002 and corresponded to an MSSA isolate from an animal in a veterinary laboratory. The earliest Scottish CC398 isolate from a human was an MSSA isolate from a blood sample collected in 2005. The isolates were from approximately equal numbers of men and women, ranging between 0 and 85 years of age. The number of antibiotics to which strains were resistant, as determined by phenotypic testing, ranged from 0 to 8 and included both MRSA and MSSA isolates (see Table S1 in the supplemental material). Seven *spa* types were identified among the sequenced Scottish CC398 isolates (t011, t034, t108, t571, t899, t1255, and t1451), and one of the sequenced isolates had a new *spa* type. There were 6 *spa* types common to both the Scottish CC398 isolates and a previously published global, non-United Kingdom CC398 data set (12).

***S. aureus* CC398 DNA extraction and whole-genome sequencing.** Seventeen of the 36 Scottish CC398 isolates were selected for whole-genome sequencing and subsequent molecular epidemiological analysis. Strains were chosen to maximize the diversity in time (isolates from 1998 to 2011) and space (with regard to the Scottish National Health Service [NHS] Health Board) (see Table S1 and Fig. S1 in the supplemental material). The stocked isolates were streaked onto blood agar plates and grown overnight at 37°C. Where strains did not grow in this first instance, an attempt was made to grow them on tryptone soy agar overnight at 37°C. One colony from each plate was selected for DNA extraction. Whole-genome sequencing, using the Illumina Mi-Seq platform, was carried out by ARK-Genomics at the Roslin Institute, Edinburgh, United Kingdom.

**Sequence assembly and alignment.** Short reads from the Illumina sequencing of CC398 isolates from Scotland were aligned against a CC398 reference sequence (GenBank accession number CP003045.1) using the Burrows-Wheeler Aligner (BWA) with the Smith-Waterman algorithm disabled (18), as described in a previous publication (19). The same approach was used to assemble publically available CC398 short reads. A core genome alignment was created from the consensus sequences, with the core genome defined as nucleotide sites shared by all sequences (alignment columns that did not contain a gap character or unknown nucleotide identity). The sequence labeled T7 in a previous publication (12) was excluded due to its conflicting phylogenetic signal.

**Recombination detection.** Previous analysis of CC398 whole-genome sequences (12) identified a region of ~123,000 bp which had been horizontally acquired from an ST9 donor; as in previous analyses (12, 20), this region of the alignment was not included in the phylogenetic analysis. After removing the ST9 donor region, the alignment of CC398 sequences was further interrogated for recombination using the BratNextGen software (21). Removal of the putative recombinant regions did not have an observable effect on the BEAST maximum clade credibility tree or on the estimates of evolutionary rates and divergence dates and did not reduce the number of variable sites by more than 45 sites (from 5,962). The consistency index for the data set, calculated in PAUP\* (22), was 0.9688 for parsimony-informative sites, indicating a low level of homoplasy.

**Preliminary phylogenetic analysis.** Maximum-likelihood phylogenies were constructed using (i) PhyML (23) and (ii) RAxML (Linux

version 7.2.8) (24). A general time-reversible nucleotide substitution model was used, with gamma-distributed rate heterogeneity across sites and 1,000 bootstrap replicates. The resulting phylogenetic trees were rooted at the midpoint between the two most divergent taxa in the tree. We investigated the relationship between root-to-tip distances and sampling dates using Path-O-Gen (<http://tree.bio.ed.ac.uk/software/pathogen/>).

**Time-scaled phylogenetic analysis.** Bayesian phylogenetic analysis was carried out using BEAST (25), with the HKY model of nucleotide substitution and gamma-distributed rate heterogeneity across sites. The uncorrelated lognormal relaxed clock was strongly preferred to the strict clock model under Bayes factor testing. For each combination of priors tested, at least two independent Markov chain Monte Carlo runs of 100 million generations were carried out, with sampling every 10,000 generations. The convergence of the runs was assessed by manually inspecting the chain traces and ensuring that the effective sample size was greater than 200 for all parameters estimated. For all runs, the first 1,000 samples were discarded as burn-in prior to further analysis. Final BEAST runs were carried out using the uncorrelated lognormal clock model with a Bayesian skyline demographic prior.

**Discrete-trait analyses.** An empirical sample of 1,000 trees was generated from the post-burn-in Bayesian skyline BEAST phylogenies. We then used discrete-trait mapping (26) to analyze the dissemination of discrete character states (host species, country of sampling, and presence or absence of antibiotic resistance genes and virulence determinants) across the phylogeny samples. “Markov jumps” analyses (27, 28) were used to count inferred transitions from one state to another (e.g., from livestock to human host or from absence to presence of a particular resistance gene).

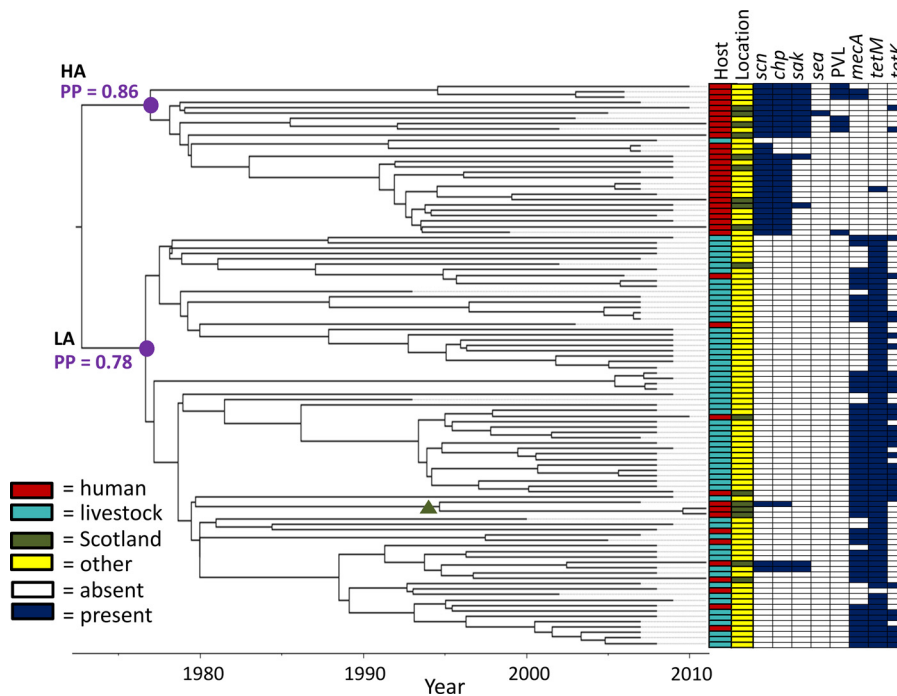
**Gene content analysis.** The presence or absence of genes carried on MGEs was investigated by mapping the short-read data for the Scottish and global CC398 isolates to reference sequences for genes of interest: the human immune evasion cluster genes *scn*, *chp*, *sak*, and *sea* (29); *mecA* and *mecC*, which confer methicillin resistance; *tetM* and *tetK*, which confer tetracycline resistance; and concatenated *lukF-PV* and *lukS-PV* for Pantone-Valentine leukocidin (PVL). Reference sequences were downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/>) or the KEGG gene database in DBGET (<http://www.genome.jp/dbget/>), with 100 nucleotides (nt) of flanking sequence at either end (see Table S2 in the supplemental material). We chose the genes to test based on a particular interest in methicillin resistance for *S. aureus* strains, the historical use of tetracyclines as growth promoters in the livestock industry, the association of PVL with community-associated *S. aureus*, and previous reports of the human immune evasion cluster genes as markers for human adaptation (29).

Only when the proportion of sites mapped to (excluding the 100-nt flanking regions) and the proportion of identity with the sites mapped to were both greater than 0.9 was a gene classified as “present.” Genes were classified as “absent” otherwise. Similar thresholds have been used in previous analyses of gene content for *S. aureus* (12), and the concordance with previously published results for CC398 strains was checked where possible. In addition, the results for presence or absence of the genes encoding toxins or antibiotic resistance were compared to the results of laboratory testing (PCR for PVL and phenotypic testing for methicillin and tetracycline resistance) for the Scottish CC398 sequences.

**Nucleotide sequence accession numbers.** The sequencing data for the 17 Scottish *S. aureus* CC398 genomes have been deposited in the European Nucleotide Archive (project accession no. PRJEB7209; sample accession numbers ERS541584, ERS541585, ERS541586, ERS541587, ERS541588, ERS541589, ERS541590, ERS541591, ERS541592, ERS541593, ERS541594, ERS541595, ERS541596, ERS541597, ERS541598, ERS541599, and ERS541600).

## RESULTS

**Distinct human- and livestock-associated CC398 clades.** We analyzed 17 *S. aureus* CC398 genomes from Scotland (see Table S1 in



**FIG 1** CC398 core genome BEAST phylogeny and distributions of hosts, locations, and genes. The time-scaled Bayesian summary phylogeny for core genome sequences revealed a split into two major clades associated with the host (human or livestock) from which the sequence was sampled. The roots of the major clades are denoted by purple dots, and posterior probability (PP) values are in purple. HA, human-associated clade; LA, livestock-associated clade. The split was also reflected in the presence or absence of genes carried on mobile genetic elements, including known markers for human adaptation (e.g., *scn* and *chp*) (34) and determinants of resistance to methicillin (*mecA*) and tetracycline (*tetM* and *tetK*). Scottish sequences from humans fell in both the human- and livestock-associated clades. A subclade (posterior probability = 1) of three Scottish human sequences within the livestock-associated clade is denoted by a green triangle.

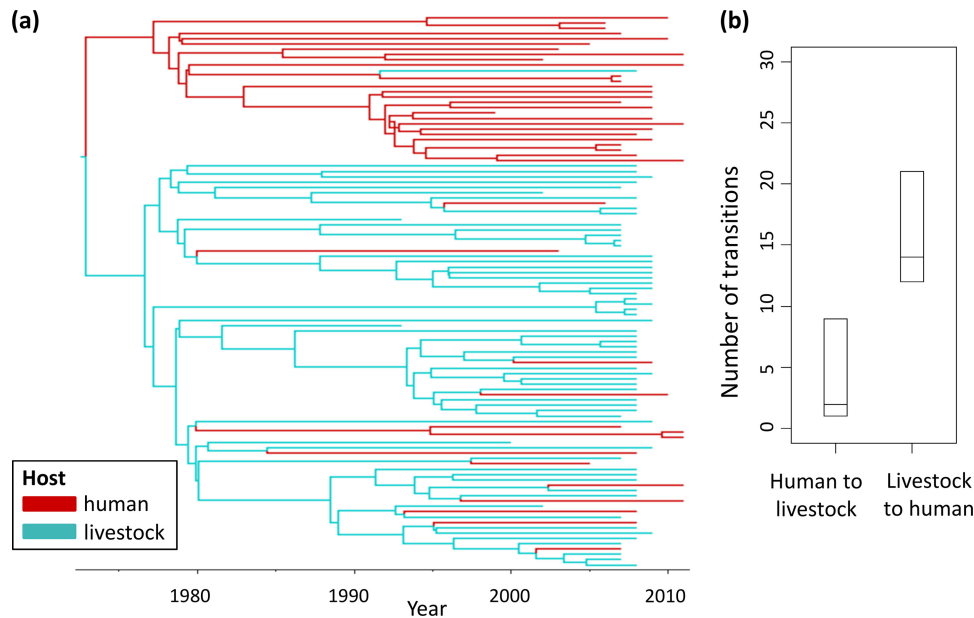
the supplemental material), along with 87 previously published CC398 genomes from humans and livestock around the world, spanning a period of 19 years (see Fig. S1 and S2 in the supplemental material), and constructed a maximum-likelihood tree from the shared (“core”) part of the genome (see Fig. S3 in the supplemental material). After rooting on the midpoint, the taxa clustered into two major clades, each with 100% bootstrap support. One clade largely consisted of human sequences (28 out of 29 members of the clade were from human hosts or environments), and the other was mainly associated with sequences of livestock origin (61 out of 76 members of the clade were from livestock hosts or environments). Scottish CC398 sequences from humans fell within both the livestock- and human-associated clades. The isolate from a Scottish veterinary laboratory fell within the livestock-associated clade. The Scottish CC398 sequences were distributed across the maximum-likelihood phylogeny (see Fig. S3 in the supplemental material), suggesting that there has been more than one introduction of CC398 into Scotland. An overall pattern of increasing root-to-tip distance with sampling date was observed (see Fig. S4 in the supplemental material).

**Time scale of the emergence of human- and livestock-associated CC398 clades.** Clustering of CC398 sequences into two major clades—one largely human associated and the other mainly livestock associated—was supported by the BEAST runs using a Bayesian skyline tree prior (Fig. 1). The human-associated clade had a posterior probability (PP) of 0.86, and the livestock-associated clade had a posterior probability of 0.78 (Fig. 1; see Fig. S5 in the supplemental material), indicating support for the separate human- and livestock-associated clades, while reflecting uncertainty in the positioning of certain sequences.

The nucleotide substitution rate was estimated at  $1.68 \times 10^{-6}$  substitutions per site per year in the BEAST analysis (95% highest posterior density [HPD] interval =  $[6.95 \times 10^{-7}, 2.49 \times 10^{-6}]$ ). This is broadly consistent with the 95% HPD interval obtained for other *S. aureus* clones (19). The most recent common ancestor (MRCA) of the CC398 phylogeny was estimated as 1966 (95% HPD interval = [1936, 1985]), and the MRCAs of the human- and livestock-associated clades were estimated as 1971 (95% HPD interval = [1945, 1987]) and 1971 (95% HPD interval = [1945, 1988]), respectively. Thus, although the human- and livestock-associated clades are estimated to have emerged around the same time, there was some uncertainty about the dates of their origin, linked to uncertainty in the rate of molecular evolution.

**Frequent livestock-to-human transmissions of CC398.** Across the evolutionary history of CC398, there were more inferred transitions from livestock hosts to humans than from humans to livestock (median = 14, 95% HPD = [12, 21] for livestock to humans; median = 2, 95% HPD = [1, 9] for humans to livestock) (Fig. 2). The 95% HPD intervals for the number of transitions between host species do not overlap, indicating that these differences are significant and implicating livestock as an important source of human CC398 infection. The root of the human-associated clade was inferred to have a human ancestral host with a high posterior probability (PP = 0.89), and the root of the livestock-associated clade was inferred to have a livestock host with a high posterior probability (PP = 1.00). There was greater uncertainty about the host state at the root of the entire tree (livestock root PP = 0.72; human root PP = 0.28), and the relatively high posterior proba-





**FIG 2** CC398 core genome BEAST phylogeny with discrete-trait mapping by host. (a) The tips of the phylogeny samples were labeled with the host (human or livestock) from which the sequence was sampled and subjected to a discrete-trait mapping analysis to infer ancestral host states. The branches are colored according to the most probable inferred ancestral host state. (b) 95% HPD intervals (upper and lower limits and median, shown on the plot as horizontal lines) were calculated for the number of transitions (Markov jumps) from humans to livestock, and livestock to humans, across the phylogeny samples.

bility assigned to a livestock root could have been overly influenced by the oldest sequences in the data set being from livestock.

**Multiple introductions of CC398 into Scotland.** The BEAST phylogeography analysis (mapping discrete country information onto the posterior phylogeny samples) indicated frequent inter-country spread of CC398 in both the human- and livestock-associated clades. Sequences from 19 countries were included in our analysis, which would require a minimum of 18 ancestral transitions between countries across the phylogeny if the sequences clustered perfectly according to country (or 36 transitions if there was complete clustering by country within each of the human- and animal-associated clades). However, on average, 79 (95% HPD = [52, 138]) transitions between countries were required to explain the distribution of locations at the tips of the phylogeny. CC398 sequences from North America, South America, and Europe were found in both the human- and livestock-associated clades, and there was evidence of clustering of sequences from Europe within the livestock-associated clade (see Fig. S6 in the supplemental material).

Introductions of CC398 into Scotland were estimated to have taken place at least 13 times (95% HPD interval = [13, 30]; median = 19) over the evolutionary history of CC398. There was a smaller number of inferred transitions from Scotland to other countries (median = 6; 95% HPD interval = [0, 20]), although there was some overlap between the 95% HPD intervals for the number of jumps in and out of Scotland. The finding that the 95% HPD interval for the number of jumps into Scotland does not include zero supports the hypothesis of multiple introductions of CC398 into Scotland.

**Historical differences in the gain and loss of resistance to methicillin and tetracycline.** Fifty-three percent of the CC398 sequences (7% of the human-associated clade and 70% of the livestock-associated clade) possessed the *mecA* gene for meth-

icillin resistance (Fig. 1) based on our *in silico* testing. The homologue *mecC* was not found among any of the CC398 isolates. Seventy-two percent of the CC398 sequences (4% of the human-associated clade and 99% of the livestock-associated clade) possessed the *tetM* gene for methicillin resistance; 43% of the *tetM*-positive isolates also carried the *tetK* tetracycline resistance gene, and two members of the human-associated clade were positive for *tetK* while being negative for *tetM*. Members of the human-associated clade were significantly more likely to have at least two immune evasion cluster genes (*scn*, *chp*, *sak*, and *sea*) associated with the prophage  $\phi$ SA3 (previously shown to be markers for human adaptation [29]) than members of the livestock-associated clade and were more likely to carry the genes for PVL, less likely to be resistant to methicillin, and less likely to carry a tetracycline resistance gene than members of the livestock-associated clade (all tested using Fisher's exact test;  $P < 0.0001$ ).

The gain and loss of genes conferring resistance to the antibiotics methicillin and tetracycline was investigated over the evolutionary history of CC398 by mapping the presence or absence of known resistance determinants as discrete traits (Table 1 and Fig. 3). Gain or loss of the *tetM* tetracycline resistance determinant occurred significantly less frequently than gain or loss of *tetK* (median number of *tetM* gain or loss events = 3, 95% HPD interval = [3, 5]; median number of *tetK* gain or loss events = 33, 95% HPD interval = [17, 110]). The median total number of gain or loss events for the methicillin resistance determinant *mecA* was 18 (95% HPD interval = [12, 29]). Almost all (94%) of the *tetK*-positive isolates were also positive for *tetM*. The overall predicted phenotypic history of tetracycline resistance was significantly more stable than for methicillin, as *tetM* was gained and lost much less frequently than *mecA*.

TABLE 1 Antibiotic resistance dynamics of CC398

Parameter <sup>a</sup>	Value		
	Median	95% HPD limit	
		Lower	Upper
<i>mecA</i>			
Total transitions	18	12	29
Gained	3	0	17
Lost	16	0	23
<i>tetM</i>			
Total transitions	3	3	5
Gained	2	1	3
Lost	2	1	3
<i>tetK</i>			
Total transitions	33	17	110
Gained	19	0	51
Lost	14	2	61

<sup>a</sup> The number of transitions from presence to absence (loss) or from absence to presence (gain) of genes conferring antibiotic resistance (*mecA* for methicillin; *tetM* or *tetK* for tetracycline) was evaluated across the BEAST phylogeny samples using a Markov jumps analysis. The total number of transitions (the sum of the number of gains and losses) for each antibiotic resistance determinant was also calculated.

## DISCUSSION

In this study, we investigated the evolutionary history of global CC398 strains in a quantitative, hypothesis-driven framework. We shed new light on the transmission and acquisition of antibiotic resistance in CC398 by (i) including CC398 genomes from the United Kingdom, (ii) creating a phylogeny with an explicit time scale to date emergence events, and (iii) analyzing the dissemination of discrete-trait data on host, country, and resistance and virulence determinants across the phylogenies.

Our analyses of CC398 sequences supported the existence of distinct human- and livestock-associated clades that emerged at similar times, with some interspecies transmission in both directions (discussed below). This finding, based upon the core genome, provides further evidence of the distinction between human and animal CC398 strains, which has previously been demonstrated for the accessory genome by microarray studies (6, 11). In accordance with these publications, we also observed a difference between the distributions of MGEs (encoding methicillin and tetracycline resistance and PVL, as well as immune evasion genes) between the human- and livestock-associated clades.

In contrast to a parsimony study by Price et al. (12), which hypothesized a human origin for CC398, we did not observe the human MSSA strains falling basal to the rest of the phylogeny; instead, they fall within the human-associated clade, with a bifurcation at the root of the tree into human- and livestock-associated clades. An advantage of our BEAST study over parsimony is that it does not rely on the choice of an outgroup sequence, which can distort a phylogeny if the outgroup is too divergent (30). Unlike parsimony, BEAST explicitly incorporates information about the sampling dates of sequences and reflects uncertainty in the phylogenetic reconstruction by considering several thousand phylogeny samples, reporting posterior probability values for clades. The discrete-trait analysis we performed did not provide strong support for choosing between a human and a livestock origin, although evidence for a potential human origin for CC398 remains from a study (10) of host distribution across a number of *S. aureus* STs.

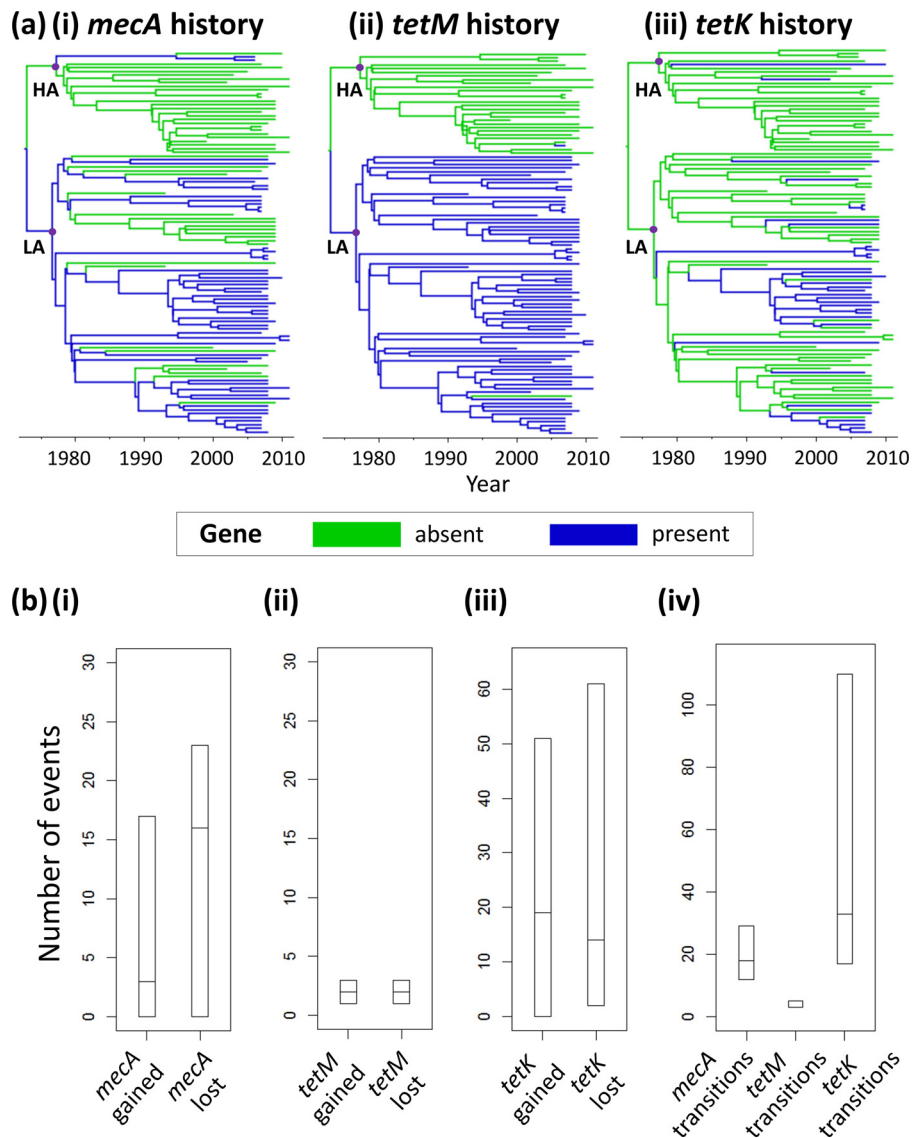
While past studies of CC398 have not included whole genomes from the United Kingdom, our inclusion of 17 previously unpublished Scottish CC398 genomes is made timely by the recent findings of CC398 in bulk milk tanks in the United Kingdom (14), as well as reports of ST398 infection in poultry in England (15). We provide evidence for multiple introductions of CC398 into Scotland, with separate introductions into the human- and livestock-associated clades. However, the source of the Scottish infections (the country and host or source from which they were transmitted) and the extent to which CC398 is able to persist and colonize within the United Kingdom remain unclear.

Of interest is a clade of three Scottish sequences in the livestock-associated clade, with 100% bootstrap support and a clade posterior probability of 1 in the maximum-likelihood and BEAST analyses, respectively (Fig. 1; see Fig. S3 in the supplemental material). The two most closely related sequences in this clade (0831N043 and 0831N044) share an ancestor around 2009 (95% HPD interval = [2006, 2010]) and were sampled approximately 6 months apart from a neonatal ward in Greater Glasgow and Clyde. The third sequence in the clade is an environmental isolate from a cleaning project in a Glasgow hospital and is more distantly related (estimated MRCA with the neonatal isolates = 1994; 95% HPD interval = [1979, 2004]). It is possible that this finding represents the persistence of the livestock-associated strain of CC398 in hospital settings in the United Kingdom, with onward transmission between humans. Such a scenario should be investigated with more intensive sampling.

We used discrete-trait mapping in BEAST to examine the phylogeography of CC398, which has been highlighted as important in the literature (31). We provided evidence for frequent transfer of CC398 between countries in both the human- and animal-associated clades. However, we did not undertake a detailed analysis of the dissemination of CC398 between individual countries, as it would likely be biased by the effects of different sampling schemes from different countries. In future, better-sampled global data sets could be analyzed with a phylogeographic generalized linear model (32) to determine the most important predictors of CC398 spread (e.g., human air travel, livestock movements, and geographical distance between locations) and to examine whether the main predictors differ between the human- and livestock-associated clades.

Phylogenetic discrete-trait mapping methods provide a state-of-the-art framework for identifying sink and source populations for infection (33–35). We provide quantitative support for previous reports that livestock are an important source of CC398 infection in humans by observing a significant number of livestock-to-human transitions within the livestock-associated clade. Coupled with the evidence for higher levels of antibiotic resistance in the livestock-associated clade, this emphasizes the need for strict biosecurity measures in agricultural settings. Humans do not appear to be a significant source of CC398 infection in livestock, although our analysis indicates that such transmission is possible.

Our analysis of antibiotic resistance dynamics across the evolutionary history of CC398 is a novel application of discrete-trait mapping methods and provides a proof of principle for a powerful new way of elucidating historical antibiotic resistance patterns in a population. For example, it allows us to determine whether resistance has arisen multiple times locally or has been acquired once and then spread widely. We noted profound differences in the histories of presence or absence of the tetracycline resistance de-



**FIG 3** CC398 antibiotic resistance histories for methicillin and tetracycline. (a) The tips of the CC398 phylogenies were labeled with presence or absence of (i) *mecA*, (ii) *tetM*, and (iii) *tetK*. The branches are colored according to the most probable inferred ancestral state (gene present or absent). HA, human-associated clade; LA, livestock-associated clade. (b) Medians and 95% HPD intervals across the phylogeny samples were plotted for the number of gains and losses of (i) *mecA*, (ii) *tetM*, and (iii) *tetK* and (iv) the total number of transitions (gains plus losses) for *mecA*, *tetM*, and *tetK*.

terminants *tetM* and *tetK* across the CC398 phylogenies. One potential explanation for this is that the *tetM* and *tetK* genes found in *S. aureus* are inherited in different ways: *tetM* is chromosomally carried on a transposon, whereas *tetK* is acquired from a plasmid (see reference 36 for details). It is possible that the mode of inheritance of the resistance determinant affects its stability, or the ability for it to be gained and lost, within a population. The *tetM* and *tetK* genes also confer resistance through different mechanisms: *tetK* through efflux (removal of the antibiotic from bacterial cells) and *tetM* through protection of the ribosomes from the action of tetracycline (see reference 36) for a discussion).

The *tetM* gene is believed to confer resistance to all antibiotics in the tetracycline group, whereas *S. aureus* strains carrying *tetK* alone are typically susceptible to some members of the tetracycline group (e.g., minocycline), and higher levels of tetracycline resistance have been found in strains harboring both *tetM* and *tetK*

(36). It is possible that there is a tradeoff between the fitness costs to the bacteria of carrying both resistance determinants and the benefit of enhanced antibiotic resistance. Discrete-trait mapping analyses on more intensively sampled data sets could help to shed light on the selective forces underlying rates of gain and loss of resistance determinants in the future. Furthermore, our *in silico* study could stimulate laboratory experiments to investigate how epistatic interactions, antibiotic usage levels, and the mode of inheritance of antibiotic resistance determinants affect their rates of gain and loss. For example, wet-laboratory studies could test whether *tetK* is lost more readily in strains that already possess *tetM* and how this relates to different levels of tetracycline usage.

The methicillin resistance determinant *mecA*, which was gained and lost repeatedly over the evolutionary history of CC398 (Fig. 3 and Table 1), particularly in the livestock-associated clade, is also carried on an MGE—the staphylococcal cassette chromo-

some *mec* element. Our analysis of gains and losses of resistance determinants revealed striking differences between the histories of methicillin and tetracycline resistance, with the methicillin resistance history being far more dynamic. Tetracycline resistance conferred by the *tetM* gene appears to have been extremely stable over the history of the livestock-associated CC398 clade, while tetracycline resistance determinants have been almost ubiquitously absent from the human-associated clade. It is possible that such differences have been driven by historical and current differences in tetracycline usage between human and animal populations. Despite reports of a decline in the use of tetracyclines in human clinical settings and a ban on their use as livestock growth promoters in Europe since the 1970s (37), tetracyclines are still routinely added to animal feed in other countries, as well as being administered in therapeutic doses on commercial farms (38).

Mapping discrete traits onto phylogenies in a Bayesian framework will provide a new way to test hypotheses about the sources and spread of bacteria, antibiotic resistance, and virulence determinants using data from natural populations. With more detailed data sets, it would be possible to formally investigate the correlation between resistance dynamics and host species or between patterns of resistance to different antibiotics. The relationship between resistance patterns and antibiotic usage levels over time could also be investigated. Genomic analysis will also allow the dynamics of resistance encoded by different mechanisms (e.g., core genome single-nucleotide polymorphisms [SNP] or different types of MGE) to be compared. Understanding the past dynamics of resistance to different antibiotics using discrete-trait analyses could help to inform future strategies for reducing resistance, such as drug cycling. In addition, the findings from discrete-trait mapping studies could be used to generate hypotheses to test through transmission experiments with animals.

In conclusion, we have performed a time-scaled phylogenetic analysis of CC398 and provided a quantitative understanding of the circulation of CC398 through separate human- and livestock-associated lineages, but with livestock also a significant source of human infection. We have also carried out a quantitative phylogenetic analysis of the loss and gain of antibiotic resistance determinants. However, human and livestock populations are linked in many ways, including agriculture, the food chain, and shared environments (39). In the future, further studies of a range of bacterial species and strains, using a large number of sequences from numerous potential sources of infection, are required to develop a more detailed understanding of the spread of bacteria and antibiotic resistance.

## ACKNOWLEDGMENTS

The project was funded by the CONtrol of COMMunity-Acquired MRSA: Rationale and Development of Counteractions Project (CONCORD) (European Commission Framework Programme 7) and by a project grant (BB/I013873/1) and institute strategic grant funding from the Biotechnology and Biological Sciences Research Council (to J.R.F.). M.J.W. was supported by the Evolution and Transfer of Antibiotic Resistance project (EvoTAR) (European Commission Framework Programme 7) and the Wellcome Trust-funded Centre for Immunity, Infection and Evolution (CIIIE) at the University of Edinburgh. C.L.G. is funded by the Scottish Universities Life Sciences Alliance and Pfizer. P.R.M. was supported by a Medical Research Council Ph.D. studentship and a Medical Research Council Centenary Award Fellowship.

We thank members of the viral phylogenetics group at the Institute of Evolutionary Biology, University of Edinburgh, for helpful discussions.

## REFERENCES

- DeLeo FR, Otto M, Kreiswirth BN, Chambers HF. 2010. Community-associated methicillin-resistant *Staphylococcus aureus*. *Lancet* 375:1557–1568. [http://dx.doi.org/10.1016/S0140-6736\(09\)61999-1](http://dx.doi.org/10.1016/S0140-6736(09)61999-1).
- Graveland H, Wagenaar JA, Bergs K, Heesterbeek H, Heederik D. 2011. Persistence of livestock associated MRSA CC398 in humans is dependent on intensity of animal contact. *PLoS One* 6:e16830. <http://dx.doi.org/10.1371/journal.pone.0016830>.
- Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* 38:1008–1015.
- Bhat M, Dulmoutier C, Taylor BS, Miller M, Vasquez G, Yunen J, Brudney K, Sanchez J, Rodriguez-Taveras C, Rojas R, Leon P, Lowy FD. 2009. *Staphylococcus aureus* ST398, New York City and Dominican Republic. *Emerg. Infect. Dis.* 15:285–287. <http://dx.doi.org/10.3201/eid1502.080609>.
- Uhlemann AC, Dumortier C, Hafer C, Taylor BS, Sanchez J, Rodriguez-Taveras C, Leon P, Rojas R, Olive C, Lowy FD. 2012. Molecular characterization of *Staphylococcus aureus* from outpatients in the Caribbean reveals the presence of pandemic clones. *Eur. J. Clin. Microbiol. Infect. Dis.* 31:505–511. <http://dx.doi.org/10.1007/s10096-011-1339-2>.
- Uhlemann AC, Porcella SF, Trivedi S, Sullivan SB, Hafer C, Kennedy AD, Barbian KD, McCarthy AJ, Street C, Hirschberg DL, Lipkin WI, Lindsay JA, DeLeo FR, Lowy FD. 2012. Identification of a highly transmissible animal-independent *Staphylococcus aureus* ST398 clone with distinct genomic and cell adhesion properties. *mBio* 3:e00027-12. <http://dx.doi.org/10.1128/mBio.00027-12>.
- Rasigade JP, Laurent F, Hubert P, Vandenesch F, Etienne J. 2010. Lethal necrotizing pneumonia caused by an ST398 *Staphylococcus aureus* strain. *Emerg. Infect. Dis.* 16:1330. <http://dx.doi.org/10.3201/eid1608.100317>.
- Fessler A, Scott C, Kadlec K, Ehrlich R, Monecke S, Schwarz S. 2010. Characterization of methicillin-resistant *Staphylococcus aureus* ST398 from cases of bovine mastitis. *J. Antimicrob. Chemother.* 65:619–625. <http://dx.doi.org/10.1093/jac/dkq021>.
- Weinert LA, Welch JJ, Suchard MA, Lemey P, Rambaut A, Fitzgerald JR. 2012. Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. *Biol. Lett.* 8:829–832. <http://dx.doi.org/10.1098/rsbl.2012.0290>.
- Shepherd MA, Fleming VM, Connor TR, Corander J, Feil EJ, Fraser C, Hanage WP. 2013. Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset. *PLoS One* 8:e62369. <http://dx.doi.org/10.1371/journal.pone.0062369>.
- McCarthy AJ, van Wamel W, Vandendriessche S, Larsen J, Denis O, Garcia-Graells C, Uhlemann AC, Lowy FD, Skov R, Lindsay JA. 2012. *Staphylococcus aureus* CC398 clade associated with human-to-human transmission. *Appl. Environ. Microbiol.* 78:8845–8848. <http://dx.doi.org/10.1128/AEM.02398-12>.
- Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdovc I, Battisti A, Franco A, Zumudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu FY, Laurent F, Keim P, Skov R, Aarestrup FM. 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 3:e00305-11. <http://dx.doi.org/10.1128/mBio.00305-11>.
- Loeffler A, Kearns AM, Ellington MJ, Smith LJ, Unt VE, Lindsay JA, Pfeiffer DU, Lloyd DH. 2009. First isolation of MRSA ST398 from UK animals: a new challenge for infection control teams? *J. Hosp. Infect.* 72:269–271. <http://dx.doi.org/10.1016/j.jhin.2009.04.002>.
- Paterson GK, Larsen J, Harrison EM, Larsen AR, Morgan FJ, Peacock SJ, Parkhill J, Zadoks RN, Holmes MA. 2012. First detection of livestock-associated methicillin-resistant *Staphylococcus aureus* CC398 in bulk tank milk in the United Kingdom, January to July 2012. *Euro Surveill.* 17:5–7.
- Department for Environment, Food, and Rural Affairs. 2013. Livestock-associated MRSA found at a farm in East Anglia. <https://www.gov.uk/government/news/livestock-associated-mrsa-found-at-a-farm-in-east-anglia>. Accessed 5 August 2014.
- European Food Safety Authority. 2009. Analysis of the baseline survey on the prevalence of methicillin-resistant *Staphylococcus aureus* (MRSA) in



- holdings with breeding pigs, in the EU, 2008. Part A: MRSA prevalence estimates. *EFSA J.* 7:1736. <http://dx.doi.org/10.2903/j.efsa.2009.1376>.
17. Dancer SJ, White LF, Lamb J, Girvan EK, Robertson C. 2009. Measuring the effect of enhanced cleaning in a UK hospital: a prospective cross-over study. *BMC Med.* 7:28. <http://dx.doi.org/10.1186/1741-7015-7-28>.
  18. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
  19. McAdam PR, Templeton KE, Edwards GF, Holden MTG, Feil EJ, Aanensen DM, Bargawi HJA, Spratt BG, Bentley SD, Parkhill J, Enright MC, Holmes A, Girvan EK, Godfrey PA, Feldgarden M, Kearns AM, Rambaut A, Robinson DA, Fitzgerald JR. 2012. Molecular tracing of the emergence, adaptation and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 109:9107–9112. <http://dx.doi.org/10.1073/pnas.1202869109>.
  20. Stegger M, Aziz M, Chroboczek T, Price LB, Ronco T, Kiil K, Skov RL, Laurent F, Andersen PS. 2013. Genome analysis of *Staphylococcus aureus* ST291, a double locus variant of ST398, reveals a distinct genetic lineage. *PLoS One* 8:e63008. <http://dx.doi.org/10.1371/journal.pone.0063008>.
  21. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, Corander J. 2012. Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids. Res.* 40:e6. <http://dx.doi.org/10.1093/nar/gkr928>.
  22. Swofford DL. 2003. Phylogenetic analysis using parsimony (\* and other methods), version 4, 4th ed. Sinauer Associates, Sunderland, MA.
  23. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321. <http://dx.doi.org/10.1093/sysbio/syq010>.
  24. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>.
  25. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973. <http://dx.doi.org/10.1093/molbev/mss075>.
  26. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. <http://dx.doi.org/10.1371/journal.pcbi.1000520>.
  27. Minin VN, Suchard MA. 2008. Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* 56:391–412. <http://dx.doi.org/10.1007/s00285-007-0120-8>.
  28. Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363:3985–3995. <http://dx.doi.org/10.1098/rstb.2008.0176>.
  29. van Wamel WJB, Rooijackers 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 beta-hemolysin-converting bacteriophages. *J. Bacteriol.* 188:1310–1315. <http://dx.doi.org/10.1128/JB.188.4.1310-1315.2006>.
  30. Qiu YL, Lee J, Whitlock BA, Bernasconi-Quadroni F, Dombrowska O. 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Mol. Biol. Evol.* 18:1745–1753. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003962>.
  31. Fitzgerald JR. 2012. Human origin for livestock-associated methicillin-resistant *Staphylococcus aureus*. *mBio* 3:e00082-12. <http://dx.doi.org/10.1128/mBio.00082-12>.
  32. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, Suchard MA. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* 10:e1003932. <http://dx.doi.org/10.1371/journal.ppat.1003932>.
  33. Goldberg TL. 2003. Application of phylogeny reconstruction and character-evolution analysis to inferring patterns of directional microbial transmission. *Prev. Vet. Med.* 61:59–70. [http://dx.doi.org/10.1016/S0167-5877\(03\)00161-2](http://dx.doi.org/10.1016/S0167-5877(03)00161-2).
  34. Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW, Petrovska L, de Pinna E, Kuroda M, Akiba M, Izumiya H, Connor TR, Suchard MA, Lemey P, Mellor DJ, Haydon DT, Thomson NR. 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 341:1514–1517. <http://dx.doi.org/10.1126/science.1240578>.
  35. Hall MD, Knowles NJ, Wadsworth J, Rambaut A, Woolhouse MEJ. 2013. Reconstructing geographical movements and host species transitions of foot-and-mouth disease virus serotype SAT 2. *mBio* 4:e00591-13. <http://dx.doi.org/10.1128/mBio.00591-13>.
  36. Trzcinski K, Cooper BS, Hryniewicz W, Dowson CG. 2000. Expression of resistance to tetracyclines in strains of methicillin-resistant *Staphylococcus aureus*. *J. Antimicrob. Chemother.* 45:763–770. <http://dx.doi.org/10.1093/jac/45.6.763>.
  37. Cogliani C, Goossens H, Greko C. 2011. Restricting antimicrobial use in food animals: lessons from Europe. *Microbe Mag.* 6:274–279. <http://dx.doi.org/10.1128/microbe.6.274.1>.
  38. Chopra I, Roberts M. 2001. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* 65:232–260. <http://dx.doi.org/10.1128/MMBR.65.2.232-260.2001>.
  39. Woolhouse MEJ, Ward MJ. 2013. Sources of antimicrobial resistance. *Science* 341:1460–1461. <http://dx.doi.org/10.1126/science.1243444>.