Whole-Genome Sequencing and Machine Learning Analysis of *Staphylococcus aureus* from Multiple Heterogeneous Sources in China Reveals Common Genetic Traits of Antimicrobial Resistance

Wei Wang,[a] Michelle Baker,[b] Yue Hu,[b] Jin Xu,[a] Dajin Yang,[a] Alexandre Maciel-Guerra,[c] Ning Xue,[b] Hui Li,[a] Shaofei Yan,[a] Menghan Li,[a] Yao Bai,[a] Yinping Dong,[a] Zixin Peng,[a] Jinjing Ma,[a,d] Fengqin Li,[a] Tania Dottorini[b]

[a]NHC Key Laboratory of Food Safety Risk Assessment, China National Center for Food Safety Risk Assessment, Beijing, China
[b]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, Leicestershire, United Kingdom
[c]School of Computer Science, University of Nottingham, Nottingham, United Kingdom
[d]School of Chemistry and Chemical Engineering, Anqing Normal University, Anqing, Anhui, China

ABSTRACT *Staphylococcus aureus* is a worldwide leading cause of numerous diseases ranging from food-poisoning to lethal infections. Methicillin-resistant *S. aureus* (MRSA) has been found capable of acquiring resistance to most antimicrobials. MRSA is ubiquitous and diverse even in terms of antimicrobial resistance (AMR) profiles, posing a challenge for treatment. Here, we present a comprehensive study of *S. aureus* in China, addressing epidemiology, phylogenetic reconstruction, genomic characterization, and identification of AMR profiles. The study analyzes 673 *S. aureus* isolates from food as well as from hospitalized and healthy individuals. The isolates have been collected over a 9-year period, between 2010 and 2018, from 27 provinces across China. By whole-genome sequencing, Bayesian divergence analysis, and supervised machine learning, we reconstructed the phylogeny of the isolates and compared them to references from other countries. We identified 72 sequence types (STs), of which, 29 were novel. We found 81 MRSA lineages by multilocus sequence type (MLST), *spa*, staphylococcal cassette chromosome *mec* element (SCC*mec*), and Panton-Valentine leukocidin (PVL) typing. In addition, novel variants of SCC*mec* type IV hosting extra metal and antimicrobial resistance genes, as well as a new SCC*mec* type, were found. New Bayesian dating of the split times of major clades showed that ST9, ST59, and ST239 in China and European countries fell in different branches, whereas this pattern was not observed for the ST398 clone. On the contrary, the clonal transmission of ST398 was more intermixed in regard to geographic origin. Finally, we identified genetic determinants of resistance to 10 antimicrobials, discriminating drug-resistant bacteria from susceptible strains in the cohort. Our results reveal the emergence of Chinese MRSA lineages enriched of AMR determinants that share similar genetic traits of antimicrobial resistance across human and food, hinting at a complex scenario of evolving transmission routes.

IMPORTANCE Little information is available on the epidemiology and characterization of *Staphylococcus aureus* in China. The role of food is a cause of major concern: staphylococcal foodborne diseases affect thousands every year, and the presence of resistant *Staphylococcus* strains on raw retail meat products is well documented. We studied a large heterogeneous data set of *S. aureus* isolates from many provinces of China, isolated from food as well as from individuals. Our large whole-genome collection represents a unique catalogue that can be easily meta-analyzed and integrated with further studies and adds to the library of *S. aureus* sequences in the public domain in a currently underrepresented geographical region. The new Bayesian dating of the split times of major

drug-resistant enriched clones is relevant in showing that Chinese and European methicillin-resistant *S. aureus* (MRSA) have evolved differently. Our machine learning approach, across a large number of antibiotics, shows novel determinants underlying resistance and reveals frequent resistant traits in specific clonal complexes, highlighting the importance of particular clonal complexes in China. Our findings substantially expand what is known of the evolution and genetic determinants of resistance in food-associated *S. aureus* in China and add crucial information for whole-genome sequencing (WGS)-based surveillance of *S. aureus*.

KEYWORDS *Staphylococcus aureus*, antimicrobial resistance, methicillin-resistant *Staphylococcus aureus* (MRSA), whole-genome sequencing, Bayesian divergence analysis, supervised machine learning, foodborne pathogen

S *taphylococcus aureus* is a leading cause of numerous illnesses ranging from food poisoning to life-threatening diseases worldwide (1–4). The bacterium has an extraordinary capacity for acquiring new antimicrobial resistance traits. Recently, the emergence and diffusion of multidrug-resistant (MDR) *S. aureus*, particularly methicillin-resistant *S. aureus* (MRSA), have attracted considerable public health attention (5–7). The World Health Organization (WHO) has included MRSA on its "global priority list of antimicrobial-resistant bacteria to guide research, discovery, and development of new antimicrobials" (8).

When first reported in 1961 (9), MRSA had been found mostly in hospitals (health care-associated MRSA [HA-MRSA]) (10, 11). By the late 1990s (12, 13), it had been found in individuals with no history of exposure to health care (community-associated MRSA [CA-MRSA]). In 2003, a livestock-associated MRSA (LA-MRSA) sequence type 398 (ST398) clone was first isolated from a human patient in Europe (14). Further evidence showed that LA-MRSA ST398 can cause infection in humans via direct contact with farm animals (15). In recent times, HA-, CA-, and LA-MRSA clones have been increasingly found outside their original reservoirs (16–20).

MRSA lineages appear to vary geographically (21). Most of the hospitals in Asia are endemic for MRSA, with two major pandemic HA-MRSA clones, ST239-staphylococcal cassette chromosome *mec* element type III/IIIA (SCC*mec* III/IIIA) and ST5-SCC*mec* II (22). Conversely, CA-MRSA varies significantly, ranging between <5% and >35% (22), and is characterized by clonal heterogeneity, with the major prevalent clones being ST30-SCC*mec* IV, ST59-SCC*mec* IV/V, ST22-SCC*mec* IV, and ST772-SCC*mec* V (20, 22). In China, ST239-SCC*mec* III/II and ST5-SCC*mec* II are also found as the prevalent HA-MRSA (22, 23), while the ST59-SCC*mec* IV/V is the prevalent clone that has been associated with CA-MRSA (22, 23). ST9 is the dominant LA-MRSA clone in China and other Asian countries as opposed to ST398 in Europe (22, 23).

S. aureus causes approximately 20% to 25% of foodborne disease outbreaks in China (24) and is included in all food surveillance programs under the Food Safety Law of the People's Republic of China. However, very limited knowledge on transmission routes and the genomic and evolutionary relationships between lineages of this bacterium has been acquired so far. Here, we report a comprehensive investigation on *S. aureus* performed on 673 isolates collected from food and human samples in China between 2010 and 2018. Epidemiology, phylogenetic reconstruction, genomic characterization, and antibacterial resistance profiling were performed using whole-genome sequencing, Bayesian divergence analysis, and several types of supervised machine learning methods.

RESULTS

High genomic diversity, novel sequence types, and MRSA and MSSA monophyletic clusters characterized by specific CCs. Food-associated *S. aureus* isolates used in this study were selected from a collection of 7,937 isolates collected between 2010 and 2018 from different foods in 27 provinces in China. For this study, we selected 593 food-associated *S. aureus* isolates (343 MRSA and 250 methicillin-susceptible *S. aureus*
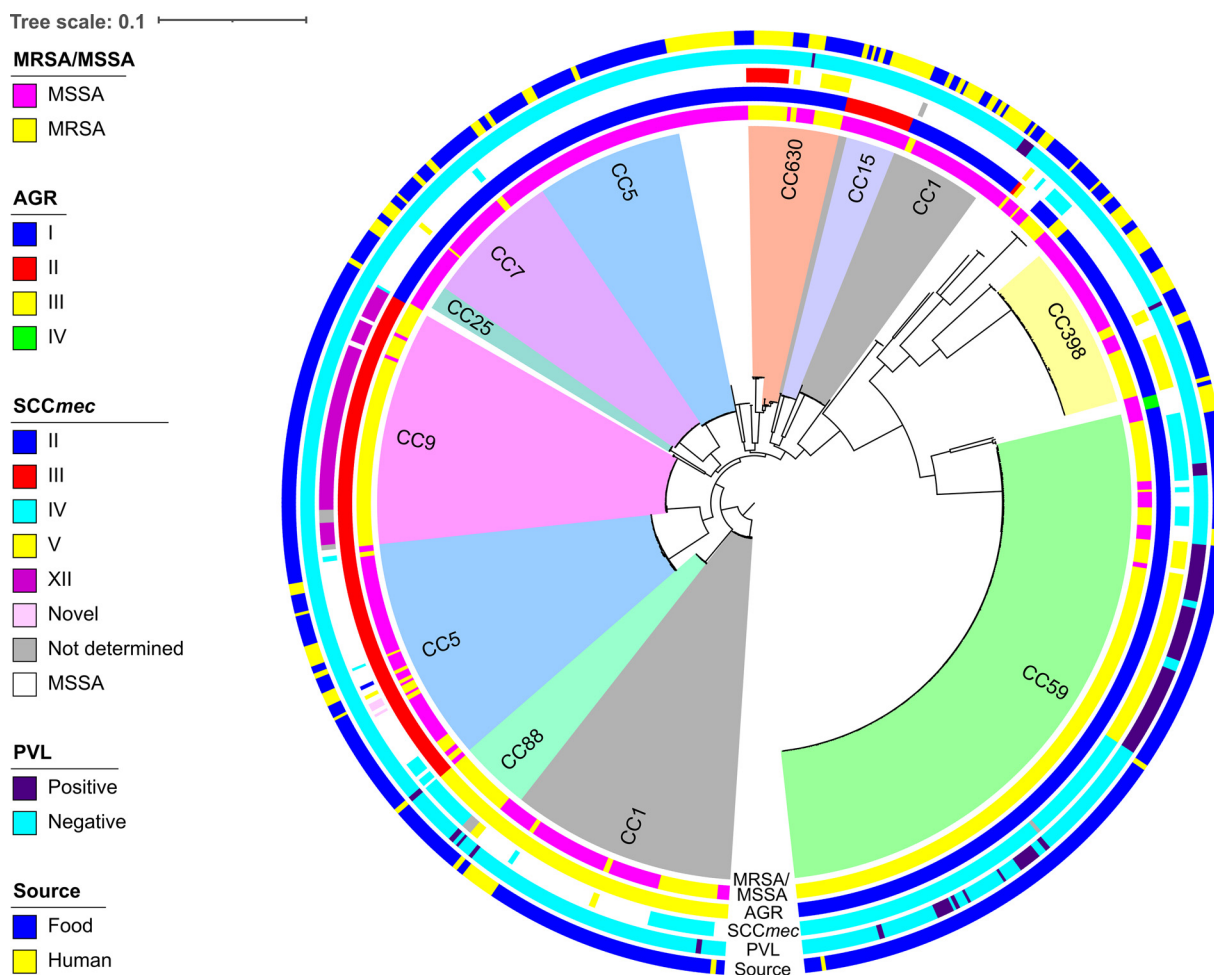
FIG 1 Sample information and phylogenetic tree of the whole cohort. Maximum likelihood phylogenetic tree based on 1,585 core genes of the 673 *S. aureus* isolates. Clonal complex clusters are distinguished by means of numbers and background colors. Methicillin-resistant and -susceptible phenotypes are indicated by the inner colored ring; *agr* subgroup, SCC*mec* type, PVL type, and the sample sources are color coded in the following rings. One branch consisting of 5 isolates was removed from the tree, as the length of the branch made visualization of the tree difficult. These isolates were all found to be *S. aureus* (subgroup *S. argenteus*). The full tree is shown in Fig. S2 in the supplemental material.

[MSSA]) and sequenced them together with 142 isolates (18 MRSA and 124 MSSA) obtained from 53 healthy and 89 infected people in Shanghai between 2015 and 2017 (see Table S1 in the supplemental material). Of these, 673 resulted in high-quality genomes, 343 being MRSA and 330 being MSSA (see Fig. S1 and Table S1). The 673 isolates were tested for susceptibility against a panel of 13 antimicrobials: 97% showed resistance to at least one, and 72% showed resistance to at least three (Fig. S1d and Table S1). As in previous studies in China (25, 26), high prevalence of resistance to penicillin (93%), erythromycin (67%), cefoxitin (58%), oxacillin (57%), and clindamycin (55%) was found. More than 99% of isolates were found to be susceptible to drugs of last resort, vancomycin and linezolid (Fig. S1d).

The phylogenetic tree was reconstructed with a set of 1,585 concatenated core genes from the 673 *S. aureus* isolates and showed strong clustering by clonal complex (CC) (27), sequence type (ST) (28), and *agr* subgroup (29) (Fig. 1). A total of 72 STs were identified, of these, 29 STs (from 47 isolates) were found to be novel (Tables S1 and 2), showing a single-base mutation in one or more alleles. Two STs (ST3656 and ST5713) of these 29 were identified from human-associated isolates. The other 27 STs were from food-associated isolates, most of which were animal food related (24 STs from 36 isolates) (Table S2). The resistance and virulence profiles of these novel STs were similar

to those of their clonal complexes. Of the 72 identified STs, 57 (representing 92% of the isolates) were related to 10 CCs, CC59 being the most prevalent (187 isolates) followed by CC5 (111 isolates). The *spa* typing (30) led to 114 different *spa* types, t437 being the most prevalent (157 isolates). The *agr* subgroups revealed that all 4 known *agr* subgroups are present in our cohort, *agr* group I being the most prevalent ($n = 418$) (Fig. 1 and Table S1).

MRSA isolates were found to cluster by CC and by type of staphylococcal cassette chromosome *mec* (SCC*mec*) (31). Interestingly, 93% of the isolates in CC59 and 97% of the isolates in CC9 were MRSA (Fig. 1). CC88 isolates were predominantly MRSA (86%) as were those of CC630 (78%) (Fig. 1). Five SCC*mec* types were identified within the 343 MRSA isolates. The predominant SCC*mec* type for food-associated MRSA was IV ($n = 165$) followed by V ($n = 87$) and XII ($n = 63$). In comparison, SCC*mec* III ($n = 11$) was dominant among human-associated MRSA isolates. Notably, SCC*mec* III and SCC*mec* XII were solely found in ST239 (CC630) and CC9, respectively, while more diversity was found within type IV and V cassettes. Analysis of the Panton-Valentine leukocidin (PVL) genes indicated that 76 isolates (11% [76/673]) were PVL positive, and 70 of these (92% [70/76]) were MRSA (Fig. 1 and Table S1). No human-associated MRSA were PVL$^+$.

The presence of different MRSA lineages was investigated by combining multilocus sequence type (MLST), *spa*, SCC*mec*, and PVL typing. We found 81 MRSA lineages across the 343 MRSA isolates (see Table S3), the five most prevalent being ST59-t437-SCC*mec* IV-PVL$^-$ ($n = 75$, 21.9%), ST9-t899-SCC*mec* XII-PVL$^-$ ($n = 53$, 15.5%), ST59-t437-SCC*mec* V-PVL$^+$ ($n = 34$, 9.9%), ST59-t437-SCC*mec* IV-PVL$^+$ ($n = 17$, 5.0%), and ST398-t011-SCC*mec* V-PVL$^-$ ($n = 13$, 3.8%). These results are consistent with those reported elsewhere in Asia (23, 32). ST59-t437-SCC*mec* IV/V isolates were primarily found in food sources, with only two isolates (both ST59-t437-SCC*mec* IV-PVL$^-$) sourced from humans. ST9-t899-SCC*mec* XII-PVL$^-$ was the predominant LA-MRSA, only found in pork, as has been seen in other studies (33, 34). Conversely, ST398-t011-SCC*mec* V-PVL$^-$ was found in different food sources, including cake, chicken, fruit, noodle, pork, sandwich, and vegetable (while in Europe and North America it is commonly found in pigs).

We searched the genomes for the presence of virulence factors and identified a wide range of genes (Table S1), including those for adherence (*clf*, *cna*, *ebp*, *fnb*, *map*, *sdr*, *spa*, *vWbp*, and *ica*), alpha-hemolysin precursor (*hly* and *hla*), aureolysin (*aur*), capsular polysaccharide (*cap8*), capsular polysaccharide (*chp*), glycerol ester hydrolase (*geh* and *scn*), hyaluronate lyase precursor (*hysA*), immune evasion (*adsA*, *esa*, *ess*, *esx*, *hlb*, *hld*, *hlg*, *lukD*, *lukF-PVL*, *lukS-PVL*, and *sbi*), ion transporter (*isd*), serine protease (*ssp*), specific sortase (*srtB*), staphylocoagulase (*coa* and *sak*), toxin (*eta*, *se*, *sel*, and *tsst-1*), and triacylglycerol lipase precursor (*lip*).

Variants of type IV SCC*mec* carrying extra antibiotic resistance genes and a novel SCC*mec* type. Among the identified 343 MRSA isolates, nearly half of the isolates (169/343) carried a type IV SCC*mec*. However, in this study, by aligning reconstructed SCC*mec* with the reference cassettes (reference [Ref] IVa, AB063172.2; Ref IVc, B096217.1), we observed several variabilities inside the type IV cassette. Two subtype IVc isolates were found to have insertions of kanamycin and bleomycin resistance genes, as previously found in Italy by Manara et al. (35). Our isolates (11A1151 and 18A25) were almost identical to the cassette (MF062) found by Manara et al. (35), see Fig. 2. Isolate 11A1151 carried a cassette identical to MF062, while isolate 18A25 had a deletion of the gene *uqpQ*. We confirmed our variant to be an integrated plasmid, pUB110 (36), via a BLAST search against N315, which is a confirmed SCC*mec* II isolate with integrated pUB110 (accession number D81934.2; identity, 99.93%). The insertion of plasmid pUB110 carrying these resistance genes is frequently associated with type II SCC*mec* elements (37–40).

An SCC*mec* IVa variant was found, in two isolates from our cohort, to have genes conferring resistance to beta-lactams; *blaI*, *blaR1*, and *blaZ* were found between IS*431* and the *mec* gene complex. When blasting this SCC*mec* variant against the NCBI nonredundant database as the subject sequence, we found that this SCC*mec* variant (99% identity and
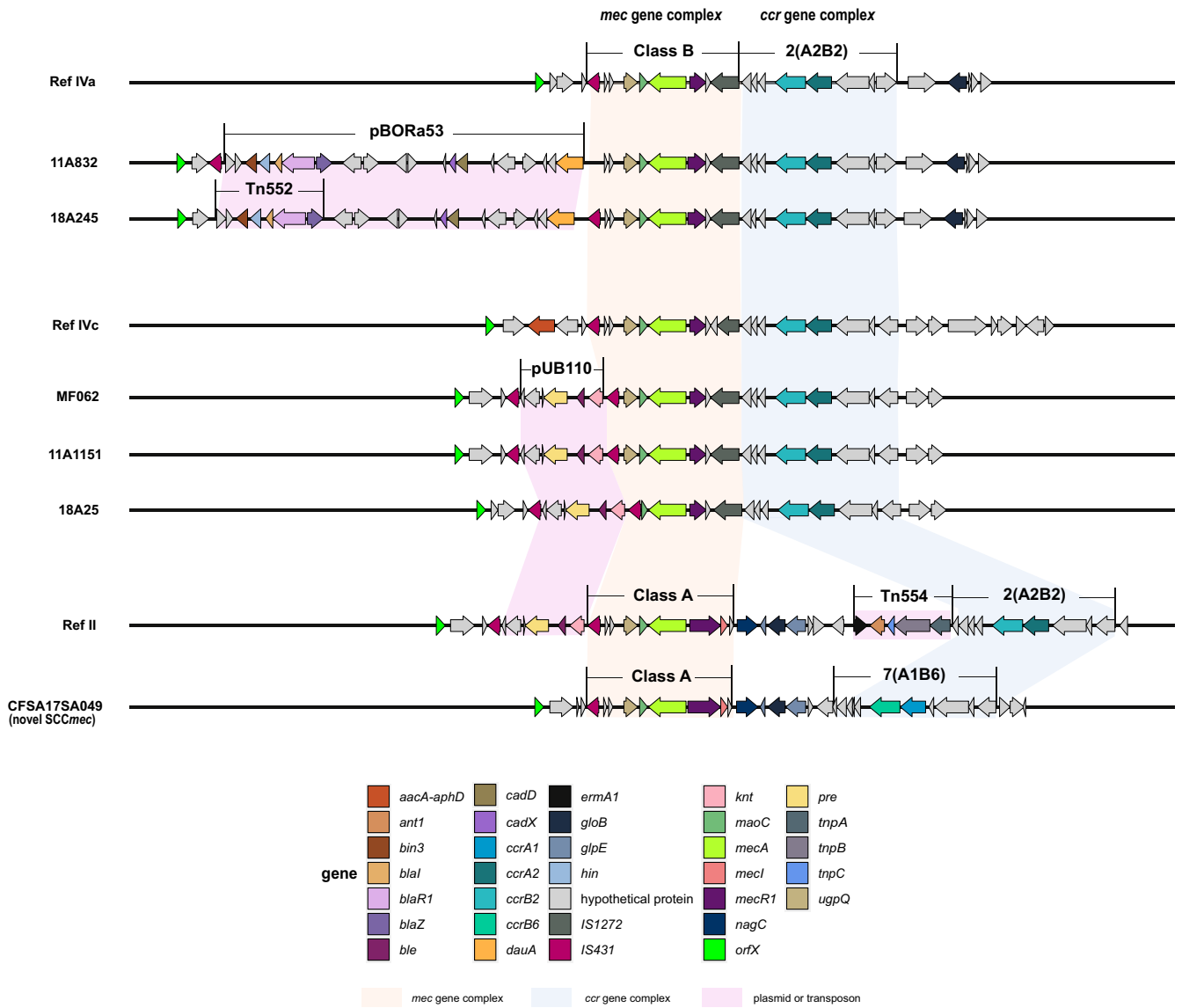
FIG 2 Schematic representation of SCC*mec* IV variability carrying extra antimicrobial resistance genes and of a novel SCC*mec* type found in our cohort. Genes are shown with the direction of transcription and color coded according to their gene name. Integrated transposons or plasmids (pBORa53, Tn*552*, Tn*554*, and pUB110) upstream or downstream of the *mec* complex and carrying resistance genes are shaded in pink. The *mec* gene complex (class A and B) and the *ccr* gene complexes 2 (A2B2) and 7 (A1B6) are shaded in light orange and blue, respectively. SCC*mec* IV elements are compared with the available references in the GenBank database (Ref IVa, AB063172.2; Ref IVc, B096217.1) for the recovered subtypes IVa and IVc. A recently reported variant of SCC*mec* IVc element (MF062, GenBank GCA_003240235) closely related to that in our study is also shown. A novel SCC*mec* element carrying the *ccr* gene complex 7 (*ccrA1* and *ccrB6*) and the *mec* gene complex class A (*mecI-mecR1-mecA*-IS*431*) is aligned to the closest reference cassette, SCC*mec* type II (Ref II, D86934.1).

query cover) was recently also found in an SCC*mec* IVa clinical isolate from Wuhan, China (GenBank CP033086.1). The genome sequence located in the J3 region (40) of the cassette indicated an integrated plasmid as its structure, where Tn*552-CadX-CadD* is associated with plasmid pBORa53 (41). The plasmid-like region also shows high similarity (100% identity and 94% query cover) to *S. aureus* plasmid pPM1 (GenBank AB699881.1).

According to the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements and recent reports, there are 14 types of SCC*mec* (31, 42–44). Three of our MRSA isolates were not attributed to any of the 14 known cassette types. These novel SCC*mec* elements carried the *ccr* gene complex 7 (A1B6) and the *mec* gene complex class A. Therefore, our study highlights the presence of further resistances and diversity within the same cassette type and the ongoing rearrangements in MRSA.

Monophyletic clustering separated STs from reference genomes of other countries. Sequence types ST59, ST9, ST398, and ST239 are widespread in China and significantly present in our cohort and include the clones most enriched in methicillin resistance. Hence, we investigated the molecular evolution and global distribution of these isolates in our cohort. The STs were analyzed to reconstruct the geographical and temporal evolution. The STs were investigated alongside publicly available isolates (PATRIC database [45]) for contextualization against the global backdrop of *S. aureus* evolution. We first reconstructed the phylogenies of the selected STs using all available good-quality reference genomes for STs, ST59 ($n = 92$), ST9 ($n = 79$), ST398 ($n = 158$), and ST239 ($n = 194$), and a core genome maximum likelihood approach (Fig. 3 and Table S4).

It is noteworthy that the Chinese ST59-SCC*mec* IV/V isolates showed monophyletic subtree profiles compared to those of the reference genomes for the same ST (Fig. 3a), suggesting a single introduction event in China followed by a spread in the Chinese territory. Exceptions were a few ST59 isolates coming from Japan ($n = 3$), Denmark ($n = 5$), and Italy ($n = 2$) that clustered with the Chinese ones, indicating potential effects of travelling and commerce. ST9 isolates from our cohort were found mainly in pork samples. They clustered with reference isolates recovered from China and a single isolate from the Netherlands (Fig. 3b and Table S4). The only exception was an isolate from pork in Beijing, which clustered with pig isolates from the United States. One possible explanation for this U.S./China clustering is that China is one of the top three markets for U.S. pork production (46), and the MRSA isolates might be transported during pork trade. As previously observed, MRSA clones can transmit cross-regionally via food trading (47–49), although evidence of a risk to humans from this is lacking. ST398 (Fig. 3c) is commonly found in livestock-associated isolates from pig farming in Europe (50, 51) and also in China (52), though at very low prevalence. It is possible that this isolate derives from ST398 sourced from swine in China. However, here, ST398 was found in higher numbers (5%, $n = 34$) from human and more diverse food sources, including pork and nonmeat products. Several studies have shown that ST398 is becoming commonly present in humans (35, 53, 54), suggesting that humans could be the most obvious source of contamination for food preparations. A total of 11 isolates (1.6% of all isolates, 78.6% of human MRSA isolates) were found to be ST239 within our cohort (Fig. 3d). Nine (with *spa* type t037) were human health care associated and clustered together, but the other two (t030) were food associated and clustered with reference Chinese human isolates (t030). The nine human isolates all carried the virulence gene *sasX*, widely associated with human-associated ST239 in China (53, 55). None of the other isolates in this study carried this gene, including the two food-associated ST239 isolates. Globally, ST239 was previously found to be associated with food animals (56, 57), but to our knowledge, this has not been found to be the case in China (58) and suggests that ST239 in China is not exclusively transmitted in clinical settings but has the potential to spread between food and humans.

To better understand the evolution of the major MRSA lineages, Bayesian divergence analysis for MRSA isolates belonging to ST59-SCC*mec* IV/V, ST9-t899-SCC*mec* XII, ST398-t011-SCC*mec* V, and ST239-t037-SCC*mec* III was performed (Fig. 4 and Table S4). Our isolates clustered away from those from other countries in the maximum likelihood trees and were characterized by different divergence times. The mean substitution rate was 2.118×10^{-6} (highest posterior density [HPD] 95%, 1.958×10^{-6} to 2.523×10^{-6}) substitutions per site per year (59, 60). For ST59-SCC*mec* IV/V, the non-Chinese isolates showed a pattern of divergence starting in the 1960s, with divergence from Chinese clones occurring in the 1940s, earlier than another study of East Asian clones (61); however, small sample sizes in both studies may explain the differences. On the contrary, Chinese isolates diversified from ~1980 onwards, with the majority of divergence occurring after ~1990, in agreement with another study of food-related *S. aureus* isolates in China (62). Our data included a small number of human samples from China within this clone (two from our cohort and three reference genomes). These isolates showed no differential evolution to the food-related *S. aureus* isolates and were most closely related to
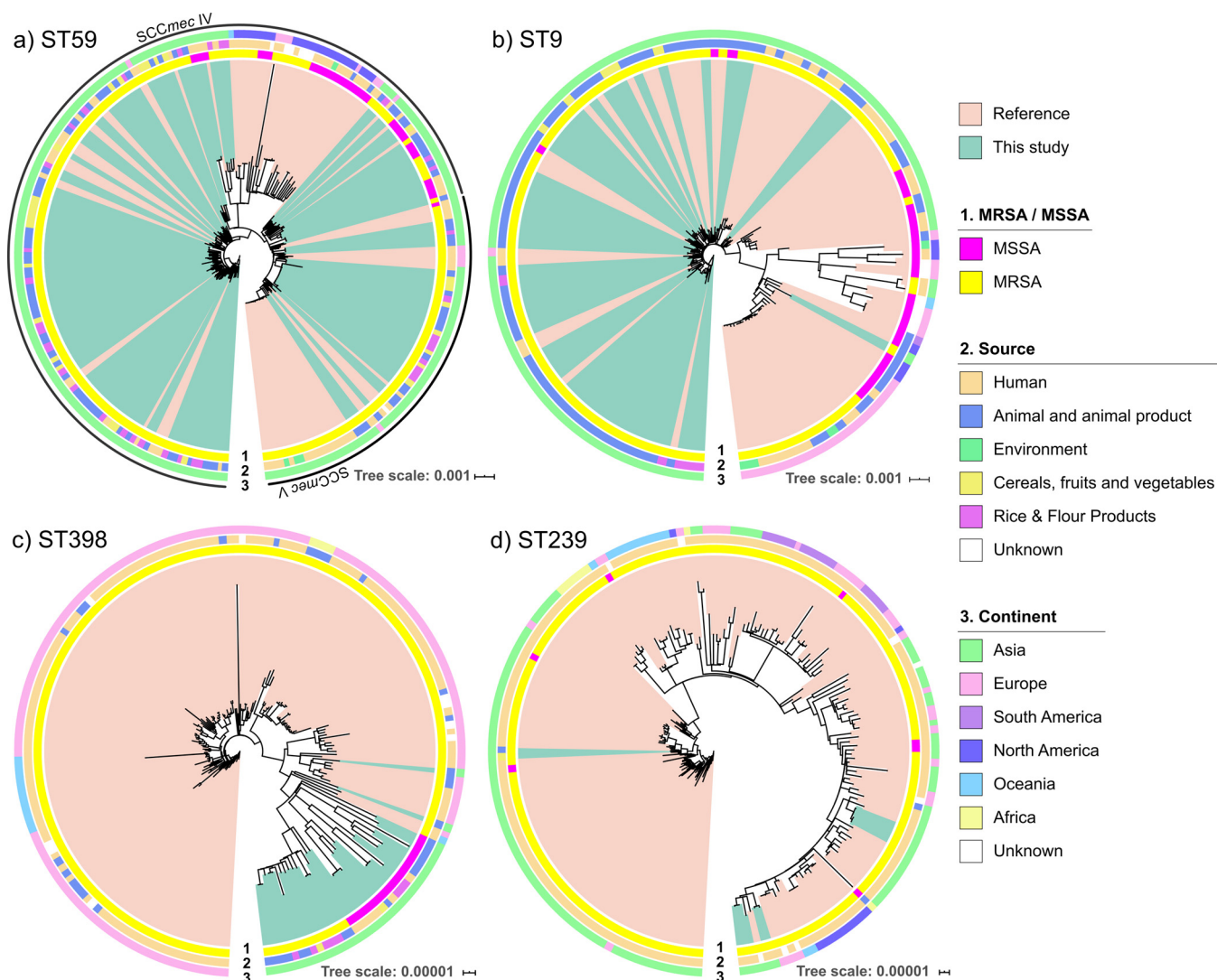
FIG 3 Whole-genome maximum likelihood phylogenetic trees of the four major ST subsets. Plots show all the isolates from this study together with all publicly available reference genomes for ST59, ST9, ST398, and ST239. Methicillin-resistant and -susceptible genotypes are indicated by the inner colored ring. Origin (sample sources) and region are color coded in the following rings. Isolates from this study and publicly available reference genomes are color coded in salmon and cyan, respectively. (a) ST59 isolates. The Chinese ST59 SCC*mec* V and ST59 SCC*mec* IV isolates clustered separately from the reference genomes collected elsewhere. ST59 isolates from our Chinese cohort were almost exclusively MRSA and *spa* type t437 and t441, in contrast to what was found elsewhere. (b) ST9 isolates. The ST9 isolates from our cohort clustered together with isolates from China and away from those collected elsewhere. ST9 isolates from our Chinese cohort were almost exclusively MRSA and *spa* type t899, in contrast to what was found elsewhere. (c) ST398 isolates. The Chinese ST398 clustered separately from the reference genomes collected elsewhere. The Chinese isolates are mainly MSSA and show a wide diversity of origin, unlike in Europe, where they are linked to pig farming. (d) ST239 isolates. The clustering indicates a close relationship between U.K. samples and samples from this study, possibly indicating international routes of transmission. Two food-associated MRSA isolates (t030) were found in this study and clustered far away from the other nine HA-MRSA isolates in this study (t037).

isolates from both meat and vegetable products, suggestive of CA (rather than HA) transmission. ST9-t899-SCC*mec* XII isolates showed much more recent diversification, with our isolates diversifying between 2004 and 2010 and the most recent common ancestor dating to the year 2000. Diversification of the Chinese ST398-t011-SCC*mec* V isolates dated back to ~2008, later than human ST398 isolates in a recent study in Taiwan (63), with their most recent common ancestor to the European strains dating back to approximately 1996. ST239-t037-SCC*mec* III showed a diverse pattern of evolution, with diversification starting in the 1950s. Our Chinese isolates diversified from 1995 onwards and were found most closely related to other samples from China. Two of our samples were more closely related to samples from Algeria with earlier diversification. The evolution of our Chinese cohort appears to differ from geographically local countries, such as Singapore (64), possibly due to different political and economic histories.
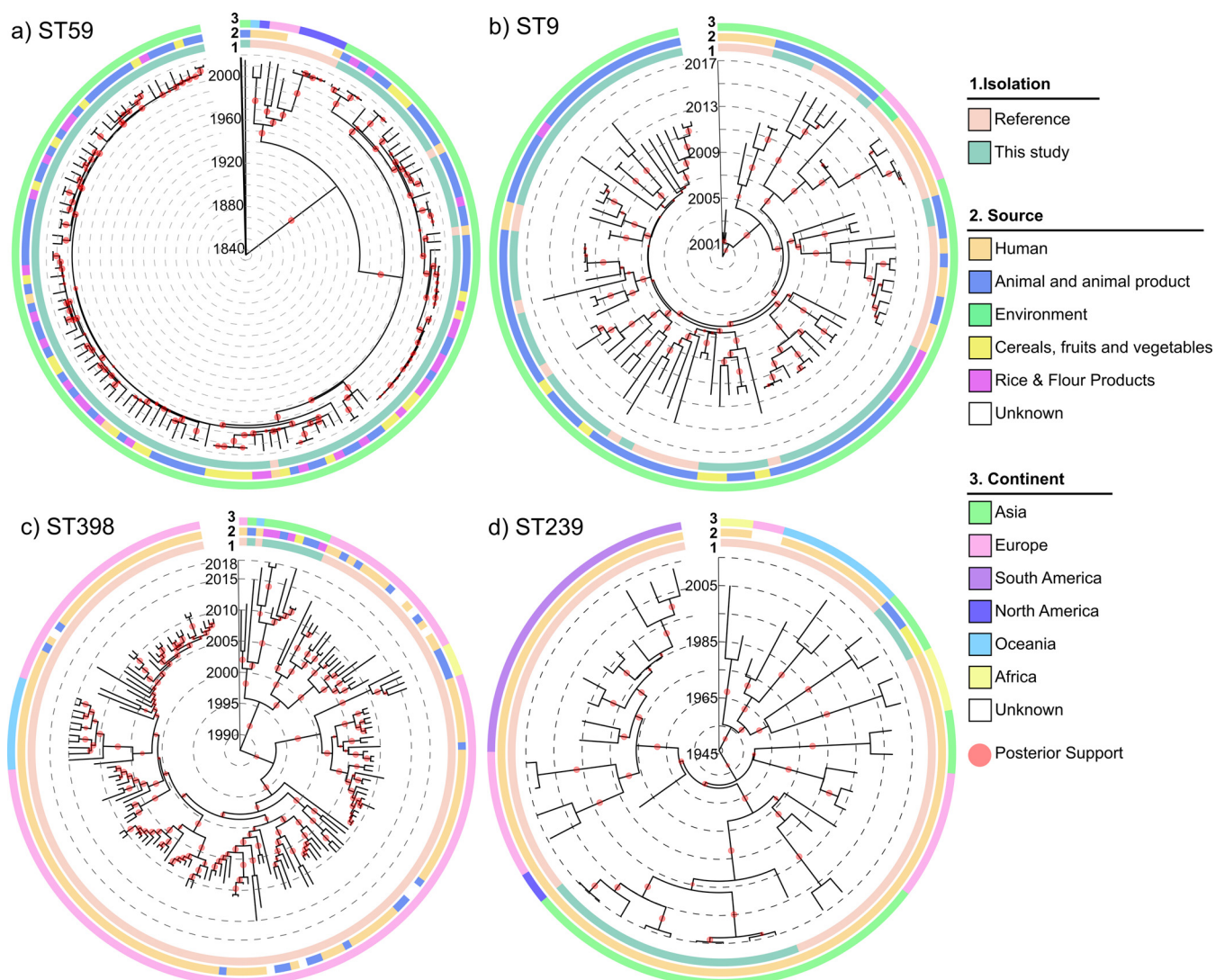
FIG 4 Evolution of four major MRSA lineages. Bayesian evolutionary analysis of ST59-SCC*mec* IV/V (a), ST9-t899-SCC*mec* XII (b), ST398-t011-SCC*mec* V (c), and ST239-t037-SCC*mec* III (d). The lineages are evolutionarily distinct. In all CC subsets, East Asian isolates cluster separately and have generally evolved later than European and American branches. Isolates from this study and publicly available reference genomes are indicated by the inner ring. Origin (sample sources) and region are color coded in the following rings. The size of the red circles on trees represents the posterior probability of each node.

Overall, the molecular evolution of MRSA in China appears to be more recent than in other countries.

Resistance to oxacillin and cefoxitin appears strongly associated with the clonal complexes CC59 and CC9. A search of features in the genome sequence of each isolate which could strongly correlate to resistance to each one of the 10 selected antimicrobials was implemented using multiple supervised machine learning methods. Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which aim to classify the data by finding an N-dimensional hyperplane to separate the data points. In this study, the best overall performance considering all antibiotic models was obtained using a linear radial basis function (RBF) SVM classifier: accuracy, 76.6% to 95.5% (range of means across all antibiotic models); area under the curve (AUC), 66.5% to 96.2% (range of means across all antibiotic models); sensitivity, 5.8% to 99.9% (range of means across all antibiotic models); specificity, 32.9% to 99.8% (range of means across all antibiotic models) (Fig. 5a and Table S5). The best predictions of resistance/susceptibility were obtained for oxacillin (accuracy, 93.85%; sensitivity, 89.42%; specificity, 99.71%; and AUC, 96.24%) and cefoxitin (accuracy, 92.83%; sensitivity, 87.70%;
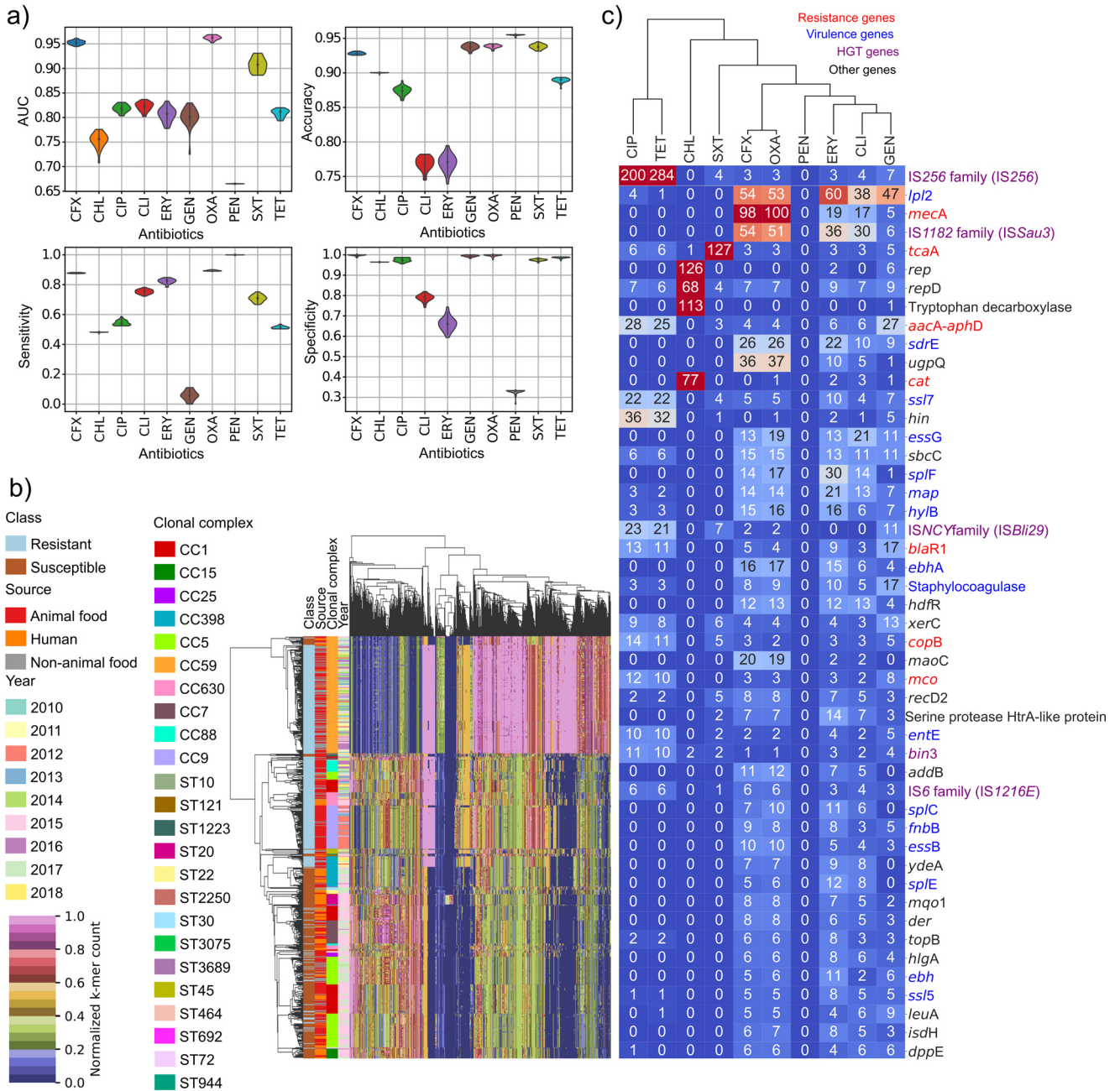
FIG 5 Supervised machine learning prediction of antimicrobial resistance signature profiles to 10 antimicrobials in the *S. aureus* Chinese cohort. (a) Prediction performance results of the RBF SVM classifier that achieved the best performance among the three investigated machine learning classifiers. The scores for each performance metric (*y* axes) are mean AUC, accuracy, sensitivity, and specificity from 30 training runs for each antimicrobial. Predictive models were generated to classify the resistance versus susceptibility profiles of 10 different antimicrobials (*x* axes). CFX, cefoxitin; CHL, chloramphenicol; CIP, ciprofloxacin; CLI, clindamycin; ERY, erythromycin; GEN, gentamicin; OXA, oxacillin; PEN, penicillin; SXT, trimethoprim-sulfamethoxazole; TET, tetracycline. (b) Hierarchical clustering of 673 isolates based on the 2,000 oxacillin-resistant genomic signatures (k-mers) recognized as the most significant by the trained classifier. Results have been data mined with respect to year of collection, ST, CC, type of sample, and source. (c) Hierarchically clustered heat map of the 10 antimicrobials based on the top 50 genes corresponding to the genomic features recognized as the most significant by the trained classifiers. Genes are color coded according to their function: resistance gene (red), virulence gene (blue), genes with function in horizontal gene transfer (HGT) (purple), genes with other functions (black). For each gene, the number of different k-mers present per gene per antibiotic model is shown on the plot, from a total of 2,000 recognized as the most significant by the trained classifier.

specificity, 99.79%; and AUC, 95.26%). Overall, 8 antibiotics achieved an AUC of >80%; chloramphenicol and penicillin did not. For oxacillin, the isolates clustered strongly by resistant/susceptible phenotype and showed strong correlation with clonal complexes despite population structure correction (Fig. 5b). Specifically, resistant isolates came from two

major clonal complexes, CC59 and CC9. CC59 was found associated with a diverse range of food sources collected over all years. Given the correlation of both oxacillin and cefoxitin to the *mecA* gene, we would expect a high degree of similarity between the features of these two models. Of the 2,000 features (k-mers) considered for each model, 1,927 (96.35%) were the same for oxacillin and cefoxitin models, and patterns of resistance seen for oxacillin were also observed in the cluster map for cefoxitin (see Data Set S1). Moreover, the cluster maps for oxacillin and cefoxitin showed a higher normalized k-mer frequency for samples belonging to CC59 than for other clonal complexes in the resistant samples. To the best of our knowledge, this difference in terms of k-mer frequency of CC59 samples was not previously observed in any other work, and it further indicates the importance of analyzing CC59 samples, especially in China. The cluster maps of the other eight antimicrobials (Data Set S1) did not show such strong clustering by susceptible/re-sistant phenotype. Notably, trimethoprim-sulfamethoxazole, tetracycline, and ciprofloxacin showed very similar patterns of resistance, with most samples being susceptible to these antimicrobials but with a small cluster of resistant samples present in all three models. The resistant isolates in this cluster were primarily associated with two clones and sample types, CC9 pork isolates and CC630 infectious human isolates.

Isolates exhibiting both trimethoprim-sulfamethoxazole and tetracycline resistance were also strongly correlated with MRSA, whereas ciprofloxacin resistance showed no association with MRSA. CC59 isolates, primarily susceptible to all three antimicrobials (trimethoprim-sulfamethoxazole, tetracycline, and ciprofloxacin) formed a large cluster on the trimethoprim-sulfamethoxazole and ciprofloxacin trees, while on the tetracy-cline tree, CC59 was more fragmented, showing that while there is clear correlation, there are important differences between these three models. Comparison of the im-portant features of the three models showed an overlap of 79.9%. Both the erythromy-cin and clindamycin models showed two separate resistant clusters, one predomi-nantly CC59 and one predominantly CC9. Other samples were also clustered by CCs but were mostly fragmented. Two antimicrobial models, gentamicin and chloramphen-icol, had very imbalanced samples, with most isolates susceptible to these antimicro-bials. In all these models, there was fragmented clustering by CC.

Machine learning reveals robust prediction of known resistance genes correlated with phenotype and novel genetic determinants of AMR phenotype. To better under-stand the relationship between antimicrobial resistance phenotype and genotype, we cross-referenced the 2,000 significant k-mers for each antibiotic model (10 of the 13 antibiotics) to the pan-genome of the 673 isolates (see Table S6) and summarized the genes found in Fig. 5c. For each antibiotic except penicillin, we obtained a list of potential genetic determinants of antibiotic resistance, with importance measured as the number of unique k-mers mapped back to the gene. For penicillin, the small number of susceptible samples impeded machine learning. For oxacillin and cefoxitin resistance, as expected, the *mecA* gene previously recognized as conferring resistance (65–67) was the primary gene found by machine learning, with 100 and 98 different k-mers, respectively, mapping back to this gene. The genes *maoC* and *ugpQ*, previously shown to be SCC*mec*-associated elements (68), were also found to be highly discriminant between resistant and susceptible cefoxitin and oxacillin phenotypes. Interestingly, *ugpQ* was also found to be in the SCC*mec*s of 342 of the 343 MRSA strains, and *maoC* was located in the SCC*mec*s of 332 isolates (see Table S7). The identification of genes known and expected to be correlated with the selected resist-ance phenotypes indicates the robustness of the methods employed, as stressed by Jaillard et al. (69). The insertion sequence IS*Sau3* (IS*1182* family) was found to be highly predictive for both antibiotics. This insertion sequence has been reported to be close to the SCC*mec* complex but has also been reported to inactivate the gene *lytH*, increasing resistance (70). Chloramphenicol resistance was highly associated with the known resistance gene *cat* (77 k-mers). The chloramphenicol resistance gene *cat* and the plasmid replication gene *rep* (also correlated with chloramphenicol resistance) were found to cooccur in 63 isolates from both animal and nonanimal food sources but were not found in any human isolates. *rep* is an ini-tiator protein in pT181 family plasmids, including pC221, known to typically carry chloram-phenicol resistance (71), indicating the likely presence of this plasmid in these food isolates.

Another gene, encoding tryptophan decarboxylase, was also significant for chloramphenicol resistance. This gene is a promising gene candidate as despite no previous known links to chloramphenicol resistance, the tryptophan biosynthesis pathway was previously associated with vancomycin (72). Interestingly, resistance to trimethoprim-sulfamethoxazole was highly associated with the gene *tcaA*. Inactivation of this gene was previously shown to increase resistance to teicoplanin and vancomycin (73), but no link has been found to trimethoprim-sulfamethoxazole. This would benefit from further experimental validation. Ciprofloxacin and tetracycline were primarily related to genes involved in horizontal gene transfer (IS*256* and IS*Bli29* [IS*NCY* family]) (Fig. 5c), suggesting these resistances may be plasmid mediated. Resistance to ciprofloxacin is typically caused by point mutation in the chromosomal *gyrA* and *parC* genes; however, there is growing evidence of plasmid-mediated resistance (74). This likely indicates a prevalence of plasmid-mediated quinolone resistance in food-related *S. aureus* isolates in China, as was previously reported in a smaller-scale study in *Escherichia coli* from farmed fish (75). Transposases, IS*Sau3* (IS*1182* family), IS*Bli29* (IS*NCY* family), and IS*256*, and the antimicrobial resistance genes *aacA-aphD* are present in many but not all isolates in our cohort from both human and animal food sources, suggesting the widespread prevalence of these genes correlated with multiple resistances in China.

The machine learning approach also revealed the presence of significant associations between virulence genes and the antimicrobial resistance profiles. Several virulence genes (e.g., *lpl2*, *essG*, *splF*, *sdrE*, *map*, and *ssl*, and others) (Table S6) were found to be associated with resistant phenotypes. Specifically, *lpl2*, a host invasion gene, was found to be a discriminant genetic feature of multiple resistant phenotypes (cefoxitin, oxacillin, ciprofloxacin, gentamicin, clindamycin, tetracycline, and erythromycin). The genes for type VII secretion (*essG*), serine protease (*splF*), serine-aspartate repeat protein (*sdrE*), and adhesin (*map*) were correlated with cefoxitin, clindamycin, erythromycin, gentamicin, and oxacillin resistance. Finally, toxin gene *ssl7* was correlated with cefoxitin, ciprofloxacin, clindamycin, erythromycin, gentamicin, oxacillin, trimethoprim-sulfamethoxazole, and tetracycline. Other virulence genes also found to be associated with antimicrobial resistance (AMR) phenotypes can be seen in Fig. 5c and Table S6.

DISCUSSION

We have analyzed a large variety of *S. aureus* samples over a 9-year temporal window collected across China (27 provinces), considering both food and human samples. In our study, we identified 29 novel sequence types (27 in food, 2 in human) with no genomic sequence available in public databases. Most *S. aureus* sequences available so far are of clinically relevant strains, and a big gap exists for less-pathogenic ones; thus, the new STs can be of relevance for further epidemiological studies leading to a better understanding of emergence, reemergence, and spread of *S. aureus* diseases. Our refined analysis of the type IV SCC*mec* highlights the presence of further resistances and plasmid insertions within this short cassette. Variants of SCC*mec* IV hosting further metal (cadmium) and antimicrobial resistance genes (kanamycin and bleomycin), and of a novel SCC*mec* type identified in this study, provide further avenues of investigation into the epidemiology of MRSA. These findings suggested that the SCC*mec* and contained resistance determinants might be transmitted via horizontal gene transfer and thus have a separate epidemiology with respect to the rest of the genomes (76, 77). Altogether, our observations shed more light on the complexity of *S. aureus* epidemiology and on the need for surveillance of the MDR and cross-host *S. aureus* to clarify the dissemination route and avoid the spread of specific genomic traits. This data set can be easily meta-analyzed and integrated with further studies, and this clearly could lead to a deeper and further understanding of the epidemiology of this bacterium and how to prevent and treat resistant infections, especially from an important area such as China.

Using a comparative genomics pipeline inspired by Manara et al. (35), we considered our isolates in the context of other similar studies. Across all isolates in our study, 97% of the *S. aureus* isolates showed resistance to at least one antimicrobial, and 72%

showed MDR, which was consistent with previous reports in China (>94% showed resistance to at least one antibiotic and >58% were MDR in food-based isolates) (78). The number of isolates carrying resistance to at least one antimicrobial found in this study was higher than in other food-based reports from Brazil (83%) (79), South Africa (71%) (80), and South Korea (51%) (81). Analogously, the rate of MDR in our study was higher than in India (53%) (82), South Korea (35%) (81), the United States (10%) (83), and Brazil (8%) (79). The recent surveillance study in Europe (53) revealed a high level of resistance among human-associated *S. aureus*, with 90% of isolates showing resistance to at least one antibiotic and 45% showing MDR, which is in accordance with our human data (84% showed resistance to at least one antibiotic and 31% were MDR).

CC59 and CC5 were the most predominant clones in this study, which agrees with those previously reported in Asia (22, 23). CC5 was also the most abundant CC type in a pan-European study of invasive *S. aureus* infections (53). Data from foodborne disease outbreaks and investigations showed that both CC59 and CC5 are common epidemic clones that cause staphylococcal food poisoning and infectious diseases (84–86). Our food isolates also showed a prevalence of SCC*mec* type IV (49.1%) and type V (25.4%), similar to previous findings in both China (78) and Germany (87). Additionally, SCC*mec* XII was also found in 63 of our food isolates. SCC*mec* XII mainly spreads as LA-MRSA in China (44) and several Asian countries, such as Japan, Malaysia, and Thailand (88–90). Recently, MRSA with SCC*mec* XII has been reported to lead to clinical infections, suggesting a potential pathogenic risk for humans (91). In our human isolates, SCC*mec* III was prevalent (9/14), which is similar to other hospital-based studies in China (92, 93), whereas SCC*mec* IV was prevalent in studies from Europe (35, 94). The *spa* types in our cohort showed a different distribution to those in another Chinese food-based study (78), with t437 most prevalent in ours compared to t071 and t091 in the study by Liao et al. (78), although all of the most prevalent isolates in that study were also present in our cohort. In an Italian study (35), *spa* types t001 (4.9%), t002 (4.3%), t008 (3.3%), and t127 (2.7%) were prevalent. While t002 (5.8%) and t127 (5.8%) were present in our human isolates, t001 and t008 were absent and *spa* types t164 (11.6%), t189 (7.2%), t037 (6.5%), and t085 (6.5%) were the most prevalent. Our cohort has a low level of PVL$^+$ human isolates (2.17%) compared to that in other studies in China (95) (12.8%), Europe (35) (27.4%), and Africa (96) (17% to 80%).

A comparison of the virulence factors present in our Chinese cohort with those of the Italian study by Manara et al. (35) shows many similarities, despite the differences in isolate hosts. Iron uptake, conserved antigen, and arginine catabolic mobile element (ACME) virulence factors found in the study by Manara et al. (35) were not present in our cohort. However, genes for immune evasion, toxins, ion transporter, adhesins, capsular polysaccharides, and serine proteases were found in similar proportions of isolates in both cohorts, though specific gene presence varied. Several immune evasion genes in our cohort are present in almost all isolates, including *esxA/B/C/D*, *hld*, and *sbi*, as found by Manara et al. (35). Additionally, the toxic shock syndrome toxin gene *tsst-1* associated with ST22 in the study by Manara et al. (35) was present in 28 of our isolates; however, in our cohort, this toxin was associated with ST1 ($n = 17$, $P < 0.0001$) and ST30 ($n = 4$, $P > 0.0001$). The presence of these virulence factors increases the risk of isolates carrying these factors harming human health, and so the widespread existence of these in food-related isolates needs to be monitored.

The CC59-t437-SCC*mec* IV/V, CC9-t899-SCC*mec* XII, and CC398-t011-SCC*mec* V clones were the most frequent food-associated MRSA clones in our study. It has been well documented that CC59-t437-SCC*mec* IV/V clones are major CA-MRSA clones in China and other Asian countries, threatening a vast population due to their epidemiological potential (22). However, they remain geographically confined and are seen in only low numbers in Europe (35, 53), suggesting importation to Europe from Asia, as was noted previously (97). The considerable numbers of this CA-MRSA lineage from food in this study could indicate some human contamination, possibly a result of inadequate hygiene measures and improper handling of food (98). Additionally, most of

the PVL-positive isolates (86.8%) belonged to CC59, demonstrating the pathogenic potential of these MRSA strains. CC9, the predominant LA-MRSA clone in Asia, was mostly identified from retail meat samples in this study. Elsewhere, CC398 is the prevalent LA-MRSA strain found in animals and humans across European countries and North America (23). Furthermore, in addition to meat, CC398-t011-SCC*mec* V MRSA isolates were detected from a greater variety of food samples, including nonanimal food such as cakes, noodles, fruits, and vegetables, suggesting a different epidemiological characteristic of these CC398 MRSA strains in China. Recently, this CC398-t011-SCC*mec* V MRSA clone was detected from humans with nonanimal contact in China and denominated as CA-MRSA (99). Thus, continued monitoring of this strain's epidemiology and preventing its widespread transmission are essential.

MRSA belonging to CC5, CC22, and CC30 clones are major HA-MRSA clones in Europe, according to a previous pan-European study (53). In our study, while all three clones are present in human samples, they are in low numbers (CC5, $n = 19$; CC22, $n = 4$; CC30, $n = 1$), and only one of these isolates was MRSA (CC5), highlighting the importance of geographical isolation in MRSA dissemination. Moreover, an Italian study on human-associated *S. aureus* in a pediatric hospital showed that CC5, CC22, CC8, CC1, and CC121 were the prevalent clones and that all CC1 and most CC5 (>96%) isolates were MRSA (35). In contrast, in our study, more than 95% of the CC1 and CC5 isolates were MSSA. In addition, for our human MRSA isolates, the major SCC*mec* type was SCC*mec* III (9/14 [64%]), while SCC*mec* IV (54/83 [67%]) was frequently detected by an Italian study and SCC*mec* III was not detected in any samples (35). Similarly, Aanensen et al. (53) only detected SCC*mec* III in clones from ST239, and the authors suggested these were likely to have been largely imported from Asia, with SCC*mec* IV again the most prevalent type in the pan-European study. However, the presence of PVL-positive isolates was strongly associated with MSSA in our study, which agreed with the Italian results (35).

CC630 has been circulating in Asia since the 1970s and is the prevalent HA-MRSA clone (22). Correspondingly, in our study, the CC630 (ST239-t037-SCC*mec* III) clone was the predominant MRSA clone among human samples (9/14), with the other isolates being ST59 (3), ST1 (1), and ST5 (1). While these are not prevalent clones in Europe, both have been found in clinical isolates. ST239 clones were present in 8 isolates in the pan-European collection (53), and ST59 was present in 3 of 86 MRSA isolates in the study by Manara et al. (35). As similarly reported by Aanensen et al. (53), the human ST239 isolates in our study carried the virulence gene *sasX*, reflecting the widespread distribution of this gene in ST239 clones in China. However, the two food-associated ST239 isolates did not carry this gene and clustered away from the other ST239 isolates in our study in the phylogenetic tree, possibly indicating that lineages not carrying this gene may be circulating in food. Additionally, other MRSA lineages, for instance, ST88-SCC*mec* IV/V, which is already known as the "African clone" (96), were also frequently detected in this study. This finding indicates that the MRSA clones have spread cross-regionally.

As in a limited number of previous phylogenetic studies, the ST9, ST59, ST239, and ST398 (58, 78) clones from our cohort tended to form separate clusters from the reference group, hinting at geographical differentiation. Meanwhile, Bayesian divergence analysis pointed to a more recent clonal evolution of MRSA in China compared to that in other countries (62, 63), potentially driven by increased economic growth and antimicrobial usage in China (100, 101). Bayesian divergence analysis also showed that while the ST9 and ST398 MRSA isolates in China exhibited relatively independent phylogenetic evolutionary relationships compared to those of isolates from other countries, there were signs of mixing of strains, possibly linked to the importation of meat products. ST239 MRSA in China showed a more recent phylogenetic relationship with isolates from other Asian countries, suggesting an introduction of ST239 MRSA to China from neighbors (64).

Several approaches to analyzing whole-genome sequences (WGS) against resistance phenotypes were previously published in the literature. These fall into two main

branches, machine learning approaches, as we have implemented here, and genome-wide association studies, a statistical association-based approach. For *S. aureus*, several groups have attempted to link resistance phenotypes to genes (53, 69, 102, 103). For example, Aanensen et al. (53) used published resistance genes, with manual curation, to predict antibiotic resistance phenotypes from genotypes with high accuracy. Although their findings provided an *in silico* typing method comparable with phenotypic accuracy, our machine-learning based approach allows impartial identification of genetic determinants of resistance, and is not limited to those already known. This is particularly important for China, where the differential evolution may have allowed resistance determinants to arise that are not as well represented in the public databases. Machine learning (ML) offers a powerful opportunity to analyze entire genomes quickly and efficiently against selected phenotypes, allowing for the identification of arbitrary numbers of sequences and other genomic features ranked on strength of correlation with the phenotype. Sequences identified by ML may contain genes with a known functional relationship with the phenotype as well as genes with no previously known association with that specific phenotype, thus providing a significant advantage to conventional bioinformatics methods based on checking for presence/absence of known manually chosen genes. Here, we have shown, for a large number of antibiotics, genes that are significantly and predictively associated with antibiotic resistance phenotypes in isolates evolved in China. Moreover, to the best of our knowledge, for the first time, we have shown that CC59 isolates have a higher k-mer frequency (for most of the studied k-mers) than other resistant clonal complexes for oxacillin and cefoxitin. This result further indicates the importance of CC59 samples, especially in China, and demonstrates the difference between this clonal complex and the other resistant clonal complexes available in our data set.

In agreement with previous studies in China (104) all our cohort isolates were susceptible to vancomycin and linezolid, while large numbers of isolates were resistant to widely used antibiotics such as penicillin, erythromycin, and cefoxitin. Thanks to ML, we were able to identify genes which, individually or in patterns, featured a strong correlation with resistance to multiple antimicrobials, regardless of the source of the isolate. As an example of the robustness of the methodology, the *maoC* and *ugpQ* genes previously found to be SCC*mec*-associated elements, in addition to *mecA*, were found strongly correlated with cefoxitin and oxacillin resistance. The identification of genes known and expected to be correlated with the selected resistant phenotypes indicates the robustness of the methods employed, as stressed by Jaillard et al. (69). ML also revealed a correlation between IS*Sau3* (IS*1182* family) and cefoxitin and oxacillin. This insertion sequence has been reported to be close to the SCC*mec* element and has also been reported to inactivate the gene *lytH*, increasing resistance (70). The *rep* gene is an initiator protein in pT181 family plasmids, including pC221, known to typically carry chloramphenicol resistance (71), indicating the likely presence of this plasmid in these food isolates. Another gene encoding tryptophan decarboxylase was also found to be significant for chloramphenicol resistance (71). Although not previously linked to chloramphenicol resistance, this gene is a promising gene candidate, as the tryptophan biosynthesis pathway has been previously associated with vancomycin resistance (72). The *tcaA* gene was found to be associated with resistance to trimethoprim-sulfamethoxazole. Inactivation of this gene was previously shown to increase resistance to teicoplanin and vancomycin, but no link has been found to trimethoprim-sulfamethoxazole. This would benefit from further experimental validation. Additionally, resistance to ciprofloxacin is typically caused by point mutations in the chromosomal *gyrA* and *parC* genes; however, there is growing evidence of plasmid-mediated resistance. In this study, the ML result showed that ciprofloxacin resistance was linked to several insertion sequences, suggesting a potential rapid spread of resistance among isolates in food and humans. Notably, using ML, significant correlations between virulence genes (*lpl2*, *essG*, *splF*, *sdrE*, *map*, and *ssl7*, plus others less strongly associated) and antimicrobial resistance phenotypes were found. The correlation between the presence of

antibiotic resistance genes (ARGs) and virulence factors was observed previously (105), and it has been proposed that an increase in virulence allows the bacteria to overcome the fitness costs associated with the carriage of AMR genes (106, 107).

Other published papers have applied machine learning to identify AMR genes associated with resistance phenotypes; however, there are many differences that make this work unique. Hyun et al. (102) employed an SVM approach to identify genes from single nucleotide polymorphisms (SNPs) based on the pan-genome of 288 *S. aureus* isolates. These isolates, all taken from human sources, were isolated in Singapore, the United States, and Russia and were primarily composed of ST239, ST22 and ST5. Hence, our study is very different; this is also highlighted by the fact that only 5 of the 13 gene candidates identified by Hyun et al. (102) (*hylX* linked to erythromycin, *oppD* and the gene for acyl coenzyme A [Acyl-CoA] linked to gentamicin, and the gene for P-type ATPase and *rep* linked to tetracycline) were also found to be important in our primarily food-based work. In addition, in our study, as described above, novel determinants underlying resistance were found strongly associated with the AMR phenotypes, such as *tcaA*, the gene for tryptophan decarboxylase, and IS*Sau3* (IS*1182* family), an original finding with respect to the literature. Both Jaillard et al. (69) and Wheeler et al. (103) used genome-wide association studies to make predictions about *S. aureus* phenotypes from the whole-genome sequences. Both studies used isolates from U.K.-based humans, giving a very different population structure to that in this study. Jaillard et al. (69) made predictions against four of the antibiotics used in our study. They also tested methicillin, which we did not; however, we tested oxacillin, which is chemically very similar and has replaced methicillin in clinical use. For these antibiotics, the predicted genes associated with the resistance phenotype for methicillin (*mecA*), erythromycin (*ermC*), and one of the 2 gentamicin genes (*aac*) were in common with our work. While two genes predicted to correlate to trimethoprim resistance in the work by Jaillard et al. (69) (*ybaK* and *mqo*1) were associated with different antibiotics in our study (*ybaK* with cefoxitin, erythromycin, gentamicin, and oxacillin and *mqo1* with cefoxitin, clindamycin, erythromycin, gentamicin, and oxacillin). This may be because, in our study, we tested trimethoprim in conjunction with sulfamethoxazole, as it is generally used clinically, rather than trimethoprim alone, as tested by Jaillard et al. (69). In the case of the study by Wheeler et al. (103), six antibiotics overlapped with our study (gentamicin, oxacillin/methicillin, erythromycin, tetracycline, ciprofloxacin, and clindamycin), and the genes found to be associated with resistance for five of these were also found by us (*aacA-aphD* linked to gentamicin, *mecA* linked to oxacillin/methicillin, *ermA* and *ermC* linked to erythromycin, *ermA* linked to clindamycin, and *tetK* and *tetM* linked to tetracycline). However, differences were found in the case of ciprofloxacin, where distinct genes were identified in the two studies (chromosomal genes *gyrA* and *parC* were significant in the study by Wheeler et al. [103], while insertion element IS*256* was found in this study). Discrepancies may reflect different resistance mechanisms, with this study indicating a prevalence of plasmid-mediated ciprofloxacin resistance in China, and the study by Wheeler et al. [103] suggesting a prevalence of chromosomal resistance mechanisms in their UK-based study.

The strong overlap between our results and previous works in identifying genes known to be correlated with the selected resistance phenotypes indicates the robustness of the methods we have employed and gives confidence to the novel predicted genes that have arisen from our analysis. Our more heterogenous data set results in a slightly lower machine learning accuracy than that in other published works, 76.6% to 95.5% compared to 91.9% to 98.6% (Hyun et al. [102]) and 94.7% to 100% (Wheeler et al. [103]). Furthermore, we have also, for the first time, found a genetic determinant of AMR for chloramphenicol and cefoxitin in *S. aureus* by using machine learning. Our study has been able to assess the differentially evolved Chinese clones of *S. aureus* and show that while many resistance mechanisms align with those seen elsewhere globally, differences may have evolved. However, we would like to point out that approaches for AMR prediction models, which have already been developed from genome sequence collections

of *S. aureus* and also for many other species, are often designed to maximize accuracy in predicting AMR phenotypes, emphasizing their diagnostic capabilities over their capacity to uncover genetic mechanisms for resistance. Many such models are also based on the detection of genes from a curated set of known AMR determinants, rendering them difficult to generalize to different treatments or organisms and unsuitable for discovering novel genes or interactions that drive resistance. In our case, we did not use known AMR determinants to search but rather the whole-genome information obtained by mapping back the k-mers on the pan-genome. Based on the concept of One Health, our study emphasizes the importance of a holistic working approach for food, human, food animals, and related sectors. The strong set of potential gene candidates identified in this work could provide new avenues of research to tackle the significant threat posed by antibiotic resistance in this part of the world.

MATERIALS AND METHODS

Sample collection and bacterial isolation. A total of 7,937 food-associated *S. aureus* isolates were cultured from various foods from 27 provinces during the years 2010 to 2018 in China. In addition, 142 *S. aureus* human-associated isolates, including 53 healthy and 89 infected human-associated isolates collected in Shanghai between 2015 and 2017, were employed in this study. All *S. aureus* isolates were confirmed using Vitek 2 Compact (bioMérieux, Craponne, France) and then were screened for MRSA by amplifying the *mecA* gene. In total, 343 food-associated isolates and 18 human-associated isolates were identified as MRSA. These MRSA isolates together with all human-associated MSSA isolates and 250 food-associated MSSA isolates (randomly selected from the full collection), giving 735 isolates in total, were used for the study of the population structure and the molecular epidemiology (see Table S1 in the supplemental material) and sent for genome sequencing. The identified isolates were stored in brain heart infusion broth with 40% (vol/vol) glycerol (HopeBio, Qingdao, China) at −80°C for the following analysis.

Antimicrobial susceptibility testing. The MICs of 13 drugs were determined using the broth dilution method by the Biofosun Gram-positive panels (Shanghai Biofosun Biotech, China) according to the CLSI guidelines (108). The panel of antimicrobial compounds tested included penicillin (PEN) (0.06 µg/ml to 8 µg/ml), oxacillin (OXA) (0.25 µg/ml to 16 µg/ml), cefoxitin (CFX) (0.25 µg/ml to 16 µg/ml), vancomycin (VAN) (0.5 µg/ml to 32 µg/ml), daptomycin (DAP) (0.125 µg/ml to 8 µg/ml), erythromycin (ERY) (0.125 µg/ml to 8 µg/ml), gentamicin (GEN) (0.5 µg/ml to 64 µg/ml), tetracycline (TET) (0.5 µg/ml-32 µg/ml), ciprofloxacin (CIP) (0.125 µg/ml to 8 µg/ml), clindamycin (CLI) (0.125 µg/ml to 8 µg/ml), trimethoprim-sulfamethoxazole (SXT) (0.125/2.3 µg/ml to 8/152 µg/ml), chloramphenicol (CHL) (1 µg/ml to 64 µg/ml), and linezolid (LZD) (0.25 µg/ml to 16 µg/ml). *S. aureus* ATCC 29213 was used as the control for the antimicrobial susceptibility testing. As no resistant breakpoint has been officially established for DAP, the resistance term (DAP-R) was assigned to strains with an MIC of >1 µg/ml in the present study to simplify the understanding (109).

DNA purification and extraction. Each isolate was grown in brain heart infusion (BHI) broth (HopeBio, Qingdao, China) at 37°C, and genomic DNA (gDNA) was purified using an Omega EZNA Bacterial DNA kit (Omega Bio-Tek, GA, USA). Genomic DNA was extracted with the sodium chloride-Tris-EDTA (STE) methods. The harvested DNA was detected by agarose gel electrophoresis and quantified by a Qubit 2.0 fluorometer (Thermo Fisher Scientific, USA).

Library construction and whole-genome sequencing. A total amount of 1 µg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext Ultra DNA library prep kit for Illumina (NEB, USA) according to the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350 bp, and then DNA fragments were end polished, A tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. Finally, PCR products were purified (AMPure XP system), and libraries were analyzed for size distribution by an Agilent 2100 Bioanalyzer and quantified using real-time PCR. The resultant DNA preps were sequenced using Illumina NovaSeq PE150 at Beijing Novogene Bioinformatics Technology Co., Ltd.

Genome assembly and annotation. All sequences were preprocessed through readfq v10 (110). To clean the data, reads containing low-quality bases (mass value ≤ 20) over 40% were removed. Those with an N in reads beyond 10% were removed. The reads which overlapped with the adapter, which exceeded 15 bp and had less than 3 mismatches between them, were removed. Clean data were processed for genome assembly with SPAdes v3.13 (111), and QUAST v4.5 (112) was used for assessing the contigs through assembly. The contigs with a length shorter than 1,000 nucleotides were filtered out. The completeness and contamination of genomes were assessed through checkM (113) with the lineage_wf pipeline. We then obtained 673 high-quality *S. aureus* genomes ($N_{50} >$ 50,000) which were used for further analysis. Genomes were annotated with Prokka v1.14.5 (114) using default parameters with -addgenes -usegenus.

***In silico* subtyping identification.** Sequence types were identified through MLST, which mapped the sequences to the PubMLST *Staphylococcus aureus* MLST database (115). Novel STs all showed one or more single base mutations from known STs and were further confirmed by PCR, according to the same protocol given on the PubMLST website (https://pubmlst.org/organisms/staphylococcus-aureus/primers),

and by DNA sequencing before being submitted to PubMLST for verification (115). Clonal complexes (CC) were annotated through the geoBURST Full MST algorithm using phyloviz software (116) with a primary founder surrounded by single-locus variants (SLVs) and known CC type in the MLST database. Spa protein A repeat region was identified through spaTyper (117) for *S. aureus* sequences. SCC*mec*Finder (118) was used for typing SCC*mec*, and the contigs that aligned to the SCC*mec* region were further annotated with Prokkav1.14.5 (114). The names of predicted proteins from Prokka were further confirmed with BLASTp (119). The variant SCC*mec* regions were further aligned using BLAST to identify SCC*mec* variants and plasmid integration into the SCC*mec* element (120, 121). SCC*mec* variants in isolates 11A1151 and 18A25 (region encompassing OrfX [*rlmH*] to the direct repeat [DR] sequence GAAGCTTATCATAAGTAA) were aligned to MF062 (GenBank GCA_003240235) and N315 (GenBank D81934.2) using default BLASTn settings (119). Pairwise comparison using BLASTn with default settings was performed for the SCC*mec* IVa variant from isolates 11A832 and 18A245 (region encompassing OrfX [*rlmH*] to the DR sequence GAAGCGTATCATAAATGA) against the NCBI nonredundant database using the SCC*mec* region and/or only the J3 region of the cassette, using default BLASTn settings.

SCC*mec* genes were visualized using the gggenes and ggplot packages in R (122). The accessory gene regulator (*agr*) subgroup was searched through BLAST using MyDbFinder 1.0 (https://cge.cbs.dtu.dk/services/MyDbFinder/) with setting parameters of -identity 90 -coverage 90. GenBank accession numbers AFS50129.1, AFS50128.1, AFS50130.1, and AFS50131.1 were used as reference sequences for *agr* type I to IV, respectively, in this study.

Virulence factors and ARG analysis. Virulence factors and ARGs were searched through Abricate software (https://github.com/tseemann/abricate) using the VFDB (123) database set B and the CARD database (124). The BLAST search within Abricate was conducted with parameters of >90% identity and >75% coverage (proportion of gene covered), in agreement with previous studies (35). Virulence gene functional categories were assigned manually based on the gene entry in the VFDB database (123). All results were manually curated by careful search of the literature. In addition, to specifically target the *fnbA/B* genes known to generate false negatives due to isoforms not present in the public databases (125), we performed an additional BLAST search, using as queries the different isoforms as described by Loughman et al. (125) and selecting the best hit entries. To identify the presence of the virulence factor *sasX* within each of our isolates, a BLASTn search was used with *sasX* (GenBank MH143577.1) as the query. Thresholds of 90% identity and 90% coverage were used.

Core gene alignment and phylogenetic analysis. All annotated files were taken as input for pangenome analysis with core gene alignments through Roary v3.13 (126). IQTree v2.0.3 (127) was then used to construct the phylogenetic trees from the core genome alignment with the general time reversible (GTR) (+F+R10) replacement model. In addition, core genome maximum likelihood phylogenetic trees clustered by different STs were also constructed with all available reference genomes downloaded from the PATRIC database (45). For ST398, only a subset of reference genomes, with collection dates, were used due to the large number of isolates available. The phylogenetic trees were subsequently visualized through iTOLv4 (128).

Bayesian divergence estimates. A subset of sequences from this study and PATRIC from ST9-t899-SCC*mec* XII, ST59-t437-SCC*mec* IV/V, ST239-t037-SCC*mec* III, and ST398-t011-SCC *mec*V in our cohort, alongside reference genomes publicly available (45) for the same lineages for which collection dates were available, were selected for Bayesian evolutionary analysis using BEAST v 1.10.4 (129). As only two reference genomes belonging to ST59-t437-SCC*mec* IV/V were available, the other 11 ST59-SCC*mec* IV/V reference genomes were also recruited to the analysis data set. Analysis was conducted on a core genome alignment of each lineage by using Roary v3.11.2 (126). All combinations of three clock models (strict, uncorrelated log normal, and uncorrelated exponential) and four tree priors (constant coalescent, logistic growth, Bayesian skyline, and birth-death model) were tested using stepping stone sampling on a subset of the isolates (ST239 isolates) to identify the best model. Log marginal likelihood values were in the range of −2,459,423 to −2,459,156. The best model was a random uncorrelated exponential clock model, with a Bayesian skyline growth model. The GTR-gamma nucleotide substitution model was used, as selected for the maximum likelihood tree. For ST59, the model did not converge after 250 million steps, and so instead, a simpler model was used (strict clock and constant coalescent growth). The analysis was run for 2 independent chains until the effective sample size (ESS), that is, the effective number of independent draws from the posterior distribution, for all parameters was greater than 200 per chain. This entailed each chain running for approximately 150 million steps. Convergence was assessed in Tracer v1.7.1 (130), and chains were subsequently combined using LogCombiner v1.10.4 (131). The maximum clade credibility tree was selected using TreeAnnotator v1.10.4 (131) and then visualized in iTOL v5 (128).

Machine learning. Machine learning was used to find features of the sample genomes that could be used to predict resistance to the panel of 13 antimicrobials. Sample genomes were first split into overlapping k-mers of 13 bp in length using GenomeTester (132) to produce a feature table for all samples. Taking each antimicrobial individually, the AMR phenotype of susceptible or resistant (Table S1) was used as the class label, with intermediate phenotypes neglected. As the classes were unbalanced, a synthetic minority oversampling technique (SMOTE) was applied to oversample data of minority class, compensating for unbalanced classes (133). The number of splits in the nested cross-validation and the number of k-nearest neighbors (default values of 5 neighbors) for SMOTE necessitated a minimum number of 12 samples in the minority class. From the panel of 13 antimicrobials, three (daptomycin, linezolid, and vancomycin) had an insufficient number of resistant isolates (minority class) to train and cross-validate the machine learning model, leaving 10 trained models. The Python package Scikit-learn (134) was used to reduce the number of features used. To correct for bias, the clonal the population structure was filtered according to k-mers based on weighted pairwise chi-squared tests between each feature and

the phenotype class, as suggested by Aun et al. 2018 (135). The weights of each genome were calculated using the method of Gerstein, Sonnhammer, and Chothia (136). Subsequently, another k-mer filtering was performed, and the top 2,000 features (k-mers) were selected with the highest chi-squared statistic (P value lower than 0.00000000000001, confidence value of more than 99.9999999%). The weighting was based on a distance matrix based on mash distances (135). A panel of machine learning algorithms was then run in the Scikit-learn package (134): logistic regression (LR), linear support vector machine (SVM), and radial basis function (RBF) SVM.

Nested cross-validation (NCV) was employed to assess the performance and select the hyperparameters of the proposed classifiers. The inner loop of the NCV found the best hyperparameters of each classifier using stratified 3-fold cross-validation; the outer loop measured the performance metrics using 5-fold stratified cross-validation. Each algorithm was run 30 times and metrics were collected for each run. The mean and standard deviation (SD) from the 30 iterations was then used as the final result statistic. The following prediction metrics were plotted using Seaborn: accuracy (true positive [TP] + true negative [TN]/[P + N]), sensitivity (true positive rate, TP/P), specificity (true negative rate: TN/N), and area under the receiver operating characteristic curve (AUC).

Identification of AMR virulence and HGT genes. Where the machine learning was able to predict the antimicrobial class based on k-mers, these were then used to search the genome for genes that contained the k-mers. Using the pan-genome of the study isolates, annotated using Prokka, the k-mers were mapped to genomes using a BLASTn query with the following parameters: evalue, 1,000; word_size, 13; gapopen, 5; gapextend, 2; outfmt, 5; strand, "plus." Genes with an identity of >70% and coverage of >70% were considered to be variants of the same gene and hence were discounted as duplicates, as done in previous literature (137); however, a more stringent threshold was used to ensure all gene variants were accounted for. The k-mer hit count (how many k-mers mapped to each identified gene) of the genes identified was then assessed for statistical significance at a significance level of 0.05 using a binomial exact test, with the probability of a gene hit based on the length of the gene and number of k-mer combinations possible per gene. All genes found were checked in the published literature to find previous associations with AMR, virulence, or horizontal gene transfer (HGT). A clustered heat map was produced using the Seaborn package in Python showing the number of k-mers mapped to each gene per antibiotic.

Data availability. Short-read sequence data for all 673 isolates used in this study are deposited in the NCBI SRA and can be found associated with BioProject PRJNA633996. The code used in this study is available in the following GitHub repository: https://github.com/tan0101/Saureus-mSytems-2021.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
DATA SET S1, XLSX file, 18.1 MB.
FIG S1, TIF file, 0.2 MB.
FIG S2, TIF file, 2.6 MB.
TABLE S1, XLSX file, 0.5 MB.
TABLE S2, XLSX file, 0.1 MB.
TABLE S3, XLSX file, 0.1 MB.
TABLE S4, XLSX file, 0.1 MB.
TABLE S5, XLSX file, 0.1 MB.
TABLE S6, XLSX file, 0.1 MB.
TABLE S7, XLSX file, 0.1 MB.

REFERENCES

1. Umeda K, Nakamura H, Yamamoto K, Nishina N, Yasufuku K, Hirai Y, Hirayama T, Goto K, Hase A, Ogasawara J. 2017. Molecular and epidemiological characterization of staphylococcal foodborne outbreak of *Staphylococcus aureus* harboring *seg*, *sei*, *sem*, *sen*, *seo*, and *selu* genes without production of classical enterotoxins. Int J Food Microbiol 256:30–35. https://doi.org/10.1016/j.ijfoodmicro.2017.05.023.

2. Hennekinne JA, De Buyser ML, Dragacci S. 2012. *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. FEMS Microbiol Rev 36:815–836. https://doi.org/10.1111/j.1574-6976.2011.00311.x.

3. Rhee Y, Aroutcheva A, Hota B, Weinstein RA, Popovich KJ. 2015. Evolving epidemiology of *Staphylococcus aureus* bacteremia. Infect Control Hosp Epidemiol 36:1417–1422. https://doi.org/10.1017/ice.2015.213.

4. Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG, Jr. 2015. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev 28:603–661. https://doi.org/10.1128/CMR.00134-14.

5. Amoako DG, Bester LA, Somboro AM, Baijnath S, Govind CN, Essack SY. 2016. Plasmid-mediated resistance and virulence mechanisms in the private health sector in KwaZulu-Natal, South Africa: an investigation of methicillin resistant *Staphylococcus aureus* (MRSA) clinical isolates collected during a three-month period. Int J Infect Dis 46:38–41. https://doi.org/10.1016/j.ijid.2016.03.019.

6. Gould IM, David MZ, Esposito S, Garau J, Lina G, Mazzei T, Peters G. 2012. New insights into methicillin-resistant *Staphylococcus aureus* (MRSA) pathogenesis, treatment and resistance. Int J Antimicrob Agents 39:96–104. https://doi.org/10.1016/j.ijantimicag.2011.09.028.

7. Rasigade JP, Dumitrescu O, Lina G. 2014. New epidemiology of *Staphylococcus aureus* infections. Clin Microbiol Infect 20:587–588. https://doi.org/10.1111/1469-0691.12718.

8. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y, Ouellette M, Outterson K, Patel J, Cavaleri M, Cox EM, Houchens CR, Grayson ML, Hansen P, Singh N, Theuretzbacher U, Magrini N, WHO Pathogens Priority List Working Group. 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis 18:318–327. https://doi.org/10.1016/S1473-3099(17)30753-3.

9. Barber M. 1961. Methicillin-resistant staphylococci. J Clin Pathol 14:385–393. https://doi.org/10.1136/jcp.14.4.385.

10. Lowy FD. 1998. *Staphylococcus aureus* infections. N Engl J Med 339:520–532. https://doi.org/10.1056/NEJM199808203390806.

11. Naimi TS, LeDell KH, Como-Sabetti K, Borchardt SM, Boxrud DJ, Etienne J, Johnson SK, Vandenesch F, Fridkin S, O'Boyle C, Danila RN, Lynfield R. 2003. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. JAMA 290:2976–2984. https://doi.org/10.1001/jama.290.22.2976.

12. Herold BC, Immergluck LC, Maranan MC, Lauderdale DS, Gaskin RE, Boyle-Vavra S, Leitch CD, Daum RS. 1998. Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. JAMA 279:593–598. https://doi.org/10.1001/jama.279.8.593.

13. Otter JA, French GL. 2010. Molecular epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in Europe. Lancet Infect Dis 10:227–239. https://doi.org/10.1016/S1473-3099(10)70053-0.

14. van Cleef BAGL, Monnet DL, Voss A, Krziwanek K, Allerberger F, Struelens M, Zemlickova H, Skov RL, Vuopio-Varkila J, Cuny C, Friedrich AW, Spiliopoulou I, Pászti J, Hardardottir H, Rossney A, Pan A, Pantosti A, Borg M, Grundmann H, Mueller-Premru M, Olsson-Liljequist B, Widmer A, Harbarth S, Schweiger A, Unal S, Kluytmans JAJW. 2011. Livestock-associated methicillin-resistant *Staphylococcus aureus* in humans, Europe. Emerg Infect Dis 17:502–505. https://doi.org/10.3201/eid1703.101036.

15. Mulders MN, Haenen APJ, Geenen PL, Vesseur PC, Poldervaart ES, Bosch T, Huijsdens XW, Hengeveld PD, Dam-Deisz WDC, Graat EAM, Mevius D, Voss A, Van De Giessen AW. 2010. Prevalence of livestock-associated MRSA in broiler flocks and risk factors for slaughterhouse personnel in the Netherlands. Epidemiol Infect 138:743–755. https://doi.org/10.1017/S0950268810000075.

16. Sieber RN, Larsen AR, Urth TR, Iversen S, Møller CH, Skov RL, Larsen J, Stegger M. 2019. Genome investigations show host adaptation and transmission of LA-MRSA CC398 from pigs into Danish healthcare institutions. Sci Rep 9:18655. https://doi.org/10.1038/s41598-019-55086-x.

17. Van Cleef BA, Verkade EJ, Wulf MW, Buiting AG, Voss A, Huijsdens XW, van Pelt W, Mulders MN, Kluytmans JA. 2010. Prevalence of livestock-associated MRSA in communities with high pig-densities in the Netherlands. PLoS One 5:e9385. https://doi.org/10.1371/journal.pone.0009385.

18. Van Den Broek I, Van Cleef B, Haenen A, Broens E, Van Der Wolf P, Van Den Broek M, Huijsdens X, Kluytmans J, Van De Giessen A, Tiemersma E. 2009. Methicillin-resistant *Staphylococcus aureus* in people living and working in pig farms. Epidemiol Infect 137:700–708. https://doi.org/10.1017/S0950268808001507.

19. Egea AL, Gagetti P, Lamberghini R, Faccone D, Lucero C, Vindel A, Tosoroni D, Garnero A, Saka HA, Galas M, *S. aureus* Study Group-Argentina, Bocco JL, Corso A, Sola C. 2014. New patterns of methicillin-resistant *Staphylococcus aureus* (MRSA) clones, community-associated MRSA genotypes behave like healthcare-associated MRSA genotypes within hospitals, Argentina. Int J Med Microbiol 304:1086–1099. https://doi.org/10.1016/j.ijmm.2014.08.002.

20. Song J-H, Hsueh P-R, Chung DR, Ko KS, Kang C-I, Peck KR, Yeom J-S, Kim S-W, Chang H-H, Kim Y-S, Jung S-I, Son JS, So T. M-k, Lalitha MK, Yang Y, Huang S-G, Wang H, Lu Q, Carlos CC, Perera JA, Chiu C-H, Liu J-W, Chongthaleong A, Thamlikitkul V, Van PH, ANSORP Study Group. 2011. Spread of methicillin-resistant *Staphylococcus aureus* between the community and the hospitals in Asian countries: an ANSORP study. J Antimicrob Chemother 66:1061–1069. https://doi.org/10.1093/jac/dkr024.

21. Chambers HF, DeLeo FR. 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. Nat Rev Microbiol 7:629–641. https://doi.org/10.1038/nrmicro2200.

22. Chen CJ, Huang YC. 2014. New epidemiology of *Staphylococcus aureus* infection in Asia. Clin Microbiol Infect 20:605–623. https://doi.org/10.1111/1469-0691.12705.

23. Chuang YY, Huang YC. 2013. Molecular epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in Asia. Lancet Infect Dis 13:698–708. https://doi.org/10.1016/S1473-3099(13)70136-1.

24. Wang X, Li G, Xia X, Yang B, Xi M, Meng J. 2014. Antimicrobial susceptibility and molecular typing of methicillin-resistant *Staphylococcus aureus* in retail foods in Shaanxi, China. Foodborne Pathog Dis 11:281–286. https://doi.org/10.1089/fpd.2013.1643.

25. Yang X, Zhang J, Yu S, Wu Q, Guo W, Huang J, Cai S. 2016. Prevalence of *Staphylococcus aureus* and methicillin-resistant *Staphylococcus aureus* in retail ready-to-eat foods in China. Front Microbiol 7:816. https://doi.org/10.3389/fmicb.2016.00816.

26. Wu S, Huang J, Zhang F, Wu Q, Zhang J, Pang R, Zeng H, Yang X, Chen M, Wang J, Dai J, Xue L, Lei T, Wei X. 2019. Prevalence and characterization of food-related methicillin-resistant *Staphylococcus aureus* (MRSA) in China. Front Microbiol 10:304. https://doi.org/10.3389/fmicb.2019.00304.

27. Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Smith JM, Murphy M, Spratt BG, Moore CE, Day NPJ. 2003. How clonal is *Staphylococcus aureus*? J Bacteriol 185:3307–3316. https://doi.org/10.1128/jb.185.11.3307-3316.2003.

28. Saunders NA, Holmes A. 2007. Multilocus sequence typing (MLST) of *Staphylococcus aureus*. Methods Mol Biol 391:71–85. https://doi.org/10.1007/978-1-59745-468-1_6.

29. Tan L, Li SR, Jiang B, Hu XM, Li S. 2018. Therapeutic targeting of the *Staphylococcus aureus* accessory gene regulator (*agr*) system. Front Microbiol 9:55. https://doi.org/10.3389/fmicb.2018.00055.

30. Koreen L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM, Kreiswirth BN. 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J Clin Microbiol 42:792–799. https://doi.org/10.1128/jcm.42.2.792-799.2004.

31. International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). 2009. Classification of staphylococcal cassette chromosome mec (SCC*mec*): guidelines for reporting novel SCC*mec* elements. Antimicrob Agents Chemother 53:4961–4967. https://doi.org/10.1128/AAC.00579-09.

32. Xiao M, Wang H, Zhao Y, Mao L-L, Brown M, Yu Y-S, O'Sullivan MVN, Kong F, Xu Y-C. 2013. National surveillance of methicillin-resistant *Staphylococcus aureus* in China highlights a still-evolving epidemiology with 15 novel emerging multilocus sequence types. J Clin Microbiol 51:3638–3644. https://doi.org/10.1128/JCM.01375-13.

33. Li J, Jiang N, Ke Y, Feßler AT, Wang Y, Schwarz S, Wu C. 2017. Characterization of pig-associated methicillin-resistant *Staphylococcus aureus*. Vet Microbiol 201:183–187. https://doi.org/10.1016/j.vetmic.2017.01.017.

34. Yan X, Li Z, Chlebowicz MA, Tao X, Ni M, Hu Y, Li Z, Grundmann H, Murray S, Pascoe B, Sheppard SK, Bo X, van Dijl JM, Du P, Zhang M, You Y, Yu X, Meng F, Wang S, Zhang J. 2016. Genetic features of livestock-

associated *Staphylococcus aureus* ST9 isolates from Chinese pigs that carry the *lsa(E)* gene for quinupristin/dalfopristin resistance. Int J Med Microbiol 306:722–729. https://doi.org/10.1016/j.ijmm.2016.08.001.

35. Manara S, Pasolli E, Dolce D, Ravenni N, Campana S, Armanini F, Asnicar F, Mengoni A, Galli L, Montagnani C, Venturini E, Rota-Stabelli O, Grandi G, Taccetti G, Segata N. 2018. Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital. Genome Med 10:82. https://doi.org/10.1186/s13073-018-0593-7.

36. McKenzie T, Hoshino T, Tanaka T, Sueoka N. 1986. The nucleotide sequence of pUB110: some salient features in relation to replication and its regulation. Plasmid 15:93–103. https://doi.org/10.1016/0147-619x(86)90046-6.

37. Ito T, Katayama Y, Hiramatsu K. 1999. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. Antimicrob Agents Chemother 43:1449–1458. https://doi.org/10.1128/AAC.43.6.1449.

38. Oliveira DC, Tomasz A, de Lencastre H. 2001. The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated *mec* elements. Microb Drug Resist 7:349–361. https://doi.org/10.1089/10766290152773365.

39. Dubin D, Matthews P, Chikramane S, Stewart P. 1991. Physical mapping of the mec region of an American methicillin-resistant *Staphylococcus aureus* strain. Antimicrob Agents Chemother 35:1661–1665. https://doi.org/10.1128/aac.35.8.1661.

40. Yamaguchi T, Ono D, Sato A. 2020. Staphylococcal cassette chromosome *mec* (SCC*mec*) analysis of MRSA, p 59–78. *In* Ji Y (ed), Methicillin-resistant Staphylococcus aureus (MRSA) protocols. Springer, Berlin, Germany.

41. Massidda O, Mingoia M, Fadda D, Whalen MB, Montanari MP, Varaldo PE. 2006. Analysis of the β-lactamase plasmid of borderline methicillin-susceptible *Staphylococcus aureus*: focus on *bla* complex genes and cadmium resistance determinants *cadD* and *cadX*. Plasmid 55:114–127. https://doi.org/10.1016/j.plasmid.2005.08.001.

42. Baig S, Johannesen TB, Overballe-Petersen S, Larsen J, Larsen AR, Stegger M. 2018. Novel SCC*mec* type XIII (9A) identified in an ST152 methicillin-resistant *Staphylococcus aureus*. Infect Genet Evol 61:74–76. https://doi.org/10.1016/j.meegid.2018.03.013.

43. Urushibara N, Aung MS, Kawaguchiya M, Kobayashi N. 2020. Novel staphylococcal cassette chromosome *mec* (SCC*mec*) type XIV (5A) and a truncated SCC*mec* element in SCC composite islands carrying *speG* in ST5 MRSA in Japan. J Antimicrob Chemother 75:46–50. https://doi.org/10.1093/jac/dkz406.

44. Wu Z, Li F, Liu D, Xue H, Zhao X. 2015. Novel type XII Staphylococcal cassette chromosome *mec* harboring a new cassette chromosome recombinase, *ccrC2*. Antimicrob Agents Chemother 59:7597–7601. https://doi.org/10.1128/AAC.01692-15.

45. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res 45:D535–D542. https://doi.org/10.1093/nar/gkw1017.

46. USDA. 2019. 2019 U.S. Agricultural export yearbook. Foreign Agricultural Service. https://www.fas.usda.gov/sites/default/files/2020-07/2019-export-yearbook.pdf. Accessed 22 May 2021.

47. Agersø Y, Hasman H, Cavaco LM, Pedersen K, Aarestrup FM. 2012. Study of methicillin resistant *Staphylococcus aureus* (MRSA) in Danish pigs at slaughter and in imported retail meat reveals a novel MRSA type in slaughter pigs. Vet Microbiol 157:246–250. https://doi.org/10.1016/j.vetmic.2011.12.023.

48. Kim YJ, Oh DH, Song BR, Heo EJ, Lim JS, Moon JS, Park HJ, Wee SH, Sung K. 2015. Molecular characterization, antibiotic resistance, and virulence factors of methicillin-resistant *Staphylococcus aureus* strains isolated from imported and domestic meat in Korea. Foodborne Pathog Dis 12:390–398. https://doi.org/10.1089/fpd.2014.1885.

49. Rodríguez-Lázaro D, Oniciuc EA, García PG, Gallego D, Fernández-Natal I, Dominguez-Gil M, Eiros-Bouza JM, Wagner M, Nicolau AI, Hernández M. 2017. Detection and characterization of *Staphylococcus aureus* and methicillin-resistant *S. aureus* in foods confiscated in EU borders. Front Microbiol 8:1344. https://doi.org/10.3389/fmicb.2017.01344.

50. Pan A, Battisti A, Zoncada A, Bernieri F, Boldini M, Franco A, Giorgi M, Iurescia M, Lorenzotti S, Martinotti M, Monaci M, Pantosti A. 2009. Community-acquired methicillin-resistant *Staphylococcus aureus* ST398 infection, Italy. Emerg Infect Dis 15:845–847. https://doi.org/10.3201/eid1505.081417.

51. Witte W, Strommenger B, Stanek C, Cuny C. 2007. Methicillin-resistant *Staphylococcus aureus* ST398 in humans and animals, central Europe. Emerg Infect Dis 13:255–258. https://doi.org/10.3201/eid1302.060924.

52. Li G, Wu C, Wang X, Meng J. 2015. Prevalence and characterization of methicillin susceptible *Staphylococcus aureus* ST398 isolates from retail foods. Int J Food Microbiol 196:94–97. https://doi.org/10.1016/j.ijfoodmicro.2014.12.002.

53. Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, Goater R, Castillo-Ramírez S, Corander J, Colijn C, Chlebowicz MA, Schouls L, Heck M, Pluister G, Ruimy R, Kahlmeter G, Åhman J, Matuschek E, Friedrich AW, Parkhill J, Bentley SD, Spratt BG, Grundmann H. 2016. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. mBio 7:e00444-16. https://doi.org/10.1128/mBio.00444-16.

54. Cuny C, Köck R, Witte W. 2013. Livestock associated MRSA (LA-MRSA) and its relevance for humans in Germany. Int J Med Microbiol 303:331–337. https://doi.org/10.1016/j.ijmm.2013.02.010.

55. Li M, Du X, Villaruz AE, Diep BA, Wang D, Song Y, Tian Y, Hu J, Yu F, Lu Y, Otto M. 2012. MRSA epidemic linked to a quickly spreading colonization and virulence determinant. Nat Med 18:816–819. https://doi.org/10.1038/nm.2692.

56. Islam MA, Parveen S, Rahman M, Huq M, Nabi A, Khan ZUM, Ahmed N, Wagenaar JA. 2019. Occurrence and characterization of methicillin resistant *Staphylococcus aureus* in processed raw foods and ready-to-eat foods in an urban setting of a developing country. Front Microbiol 10:503. https://doi.org/10.3389/fmicb.2019.00503.

57. Nemeghaire S, Roelandt S, Argudín MA, Haesebrouck F, Butaye P. 2013. Characterization of methicillin-resistant *Staphylococcus aureus* from healthy carrier chickens. Avian Pathol 42:342–346. https://doi.org/10.1080/03079457.2013.805183.

58. Liao F, Mo Z, Gu W, Xu W, Fu X, Zhang Y. 2020. A comparative genomic analysis between methicillin-resistant *Staphylococcus aureus* strains of hospital acquired and community infections in Yunnan province of China. BMC Infect Dis 20:137. https://doi.org/10.1186/s12879-020-4866-6.

59. 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. https://doi.org/10.1073/pnas.1202869109.

60. Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Žemličková H, Leblois R, Wirth T, Jombart T, Balloux F, Witte W. 2010. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. PLoS Pathog 6:e1000855. https://doi.org/10.1371/journal.ppat.1000855.

61. Ward MJ, Goncheva M, Richardson E, McAdam PR, Raftis E, Kearns A, Daum RS, David MZ, Lauderdale TL, Edwards GF, Nimmo GR, Coombs GW, Huijsdens X, Woolhouse MEJ, Fitzgerald JR. 2016. Identification of source and sink populations for the emergence and global spread of the East-Asia clone of community-associated MRSA. Genome Biol 17:160–160. https://doi.org/10.1186/s13059-016-1022-0.

62. Pang R, Wu S, Zhang F, Huang J, Wu H, Zhang J, Li Y, Ding Y, Zhang J, Chen M, Wei X, Zhang Y, Gu Q, Zhou Z, Liang B, Li W, Wu Q. 2020. The genomic context for the evolution and transmission of community-associated *Staphylococcus aureus* ST59 through the food chain. Front Microbiol 11:422. https://doi.org/10.3389/fmicb.2020.00422.

63. Huang Y-C, Chen C-J. 2020. Detection and phylogeny of *Staphylococcus aureus* sequence type 398 in Taiwan. J Biomed Sci 27:15. https://doi.org/10.1186/s12929-019-0608-8.

64. Hsu L-Y, Harris SR, Chlebowicz MA, Lindsay JA, Koh T-H, Krishnan P, Tan T-Y, Hon P-Y, Grubb WB, Bentley SD, Parkhill J, Peacock SJ, Holden MTG. 2015. Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. Genome Biol 16:81–81. https://doi.org/10.1186/s13059-015-0643-z.

65. Broekema NM, Van TT, Monson TA, Marshall SA, Warshauer DM. 2009. Comparison of cefoxitin and oxacillin disk diffusion methods for detection of *mecA*-mediated resistance in *Staphylococcus aureus* in a large-scale study. J Clin Microbiol 47:217–219. https://doi.org/10.1128/JCM.01506-08.

66. Hososaka Y, Hanaki H, Endo H, Suzuki Y, Nagasawa Z, Otsuka Y, Nakae T, Sunakawa K. 2007. Characterization of oxacillin-susceptible *mecA*-positive *Staphylococcus aureus*: a new type of MRSA. J Infect Chemother 13:79–86. https://doi.org/10.1007/s10156-006-0502-7.

67. Swenson JM, Tenover FC, Cefoxitin Disk Study Group. 2005. Results of disk diffusion testing with cefoxitin correlate with presence of *mecA* in *Staphylococcus* spp. J Clin Microbiol 43:3818–3823. https://doi.org/10.1128/JCM.43.8.3818-3823.2005.

68. Monecke S, Jatzwauk L, Müller E, Nitschke H, Pfohl K, Slickers P, Reissig A, Ruppelt-Lorz A, Ehricht R. 2016. Diversity of SCC*mec* elements in *Staphylococcus aureus* as observed in south-eastern Germany. PLoS One 11:e0162654. https://doi.org/10.1371/journal.pone.0162654.

69. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, Jacob L. 2018. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events. PLoS Genet 14:e1007758. https://doi.org/10.1371/journal.pgen.1007758.

70. Fujimura T, Murakami K. 2008. *Staphylococcus aureus* clinical isolate with high-level methicillin resistance with an *lytH* mutation caused by IS*1182* insertion. Antimicrob Agents Chemother 52:643–647. https://doi.org/10.1128/AAC.00395-07.

71. Kwong SM, Ramsay JP, Jensen SO, Firth N. 2017. Replication of staphylococcal resistance plasmids. Front Microbiol 8:2279–2279. https://doi.org/10.3389/fmicb.2017.02279.

72. Matsuo M, Cui L, Kim J, Hiramatsu K. 2013. Comprehensive identification of mutations responsible for heterogeneous vancomycin-intermediate *Staphylococcus aureus* (hVISA)-to-VISA conversion in laboratory-generated VISA strains derived from hVISA clinical strain Mu3. Antimicrob Agents Chemother 57:5843–5853. https://doi.org/10.1128/AAC.00425-13.

73. Maki H, McCallum N, Bischoff M, Wada A, Berger-Bächi B. 2004. *tcaA* inactivation increases glycopeptide resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother 48:1953–1959. https://doi.org/10.1128/AAC.48.6.1953-1959.2004.

74. Kim HB, Park CH, Kim CJ, Kim E-C, Jacoby GA, Hooper DC. 2009. Prevalence of plasmid-mediated quinolone resistance determinants over a 9-year period. Antimicrob Agents Chemother 53:639–645. https://doi.org/10.1128/AAC.01051-08.

75. Jiang HX, Tang D, Liu YH, Zhang XH, Zeng ZL, Xu L, Hawkey PM. 2012. Prevalence and characteristics of β-lactamase and plasmid-mediated quinolone resistance genes in *Escherichia coli* isolated from farmed fish in China. J Antimicrob Chemother 67:2350–2353. https://doi.org/10.1093/jac/dks250.

76. Archer GL, Niemeyer DM. 1994. Origin and evolution of DNA associated with resistance to methicillin in staphylococci. Trends Microbiol 2:343–347. https://doi.org/10.1016/0966-842x(94)90608-4.

77. Hiramatsu K. 2001. Vancomycin-resistant *Staphylococcus aureus*: a new model of antibiotic resistance. Lancet Infect Dis 1:147–155. https://doi.org/10.1016/S1473-3099(01)00091-3.

78. Liao F, Gu W, Yang Z, Mo Z, Fan L, Guo Y, Fu X, Xu W, Li C, Dai J. 2018. Molecular characteristics of *Staphylococcus aureus* isolates from food surveillance in southwest China. BMC Microbiol 18:91. https://doi.org/10.1186/s12866-018-1239-z.

79. Kroning IS, Iglesias MA, Sehn CP, Valente Gandra TK, Mata MM, da Silva WP. 2016. *Staphylococcus aureus* isolated from handmade sweets: biofilm formation, enterotoxigenicity and antimicrobial resistance. Food Microbiol 58:105–111. https://doi.org/10.1016/j.fm.2016.04.001.

80. Pekana A, Green E. 2018. Antimicrobial resistance profiles of *Staphylococcus aureus* isolated from meat carcasses and bovine milk in abattoirs and dairy farms of the Eastern Cape, South Africa. Int J Environ Res Public Health 15:2223. https://doi.org/10.3390/ijerph15102223.

81. Kim YB, Seo KW, Jeon HY, Lim SK, Lee YJ. 2018. Characteristics of the antimicrobial resistance of *Staphylococcus aureus* isolated from chicken meat produced by different integrated broiler operations in Korea. Poult Sci 97:962–969. https://doi.org/10.3382/ps/pex357.

82. Zehra A, Gulzar M, Singh R, Kaur S, Gill JPS. 2019. Prevalence, multidrug resistance and molecular typing of methicillin-resistant *Staphylococcus aureus* (MRSA) in retail meat from Punjab, India. J Glob Antimicrob Resist 16:152–158. https://doi.org/10.1016/j.jgar.2018.10.005.

83. Ge B, Mukherjee S, Hsu CH, Davis JA, Tran TTT, Yang Q, Abbott JW, Ayers SL, Young SR, Crarey ET, Womack NA, Zhao S, McDermott PF. 2017. MRSA and multidrug-resistant *Staphylococcus aureus* in U.S. retail meats, 2010-2011. Food Microbiol 62:289–297. https://doi.org/10.1016/j.fm.2016.10.029.

84. Li S, Sun S, Yang C, Chen H, Yin Y, Li H, Zhao C, Wang H. 2018. The changing pattern of population structure of *Staphylococcus aureus* from bacteremia in China from 2013 to 2016: ST239-030-MRSA replaced by ST59-t437. Front Microbiol 9:332. https://doi.org/10.3389/fmicb.2018.00332.

85. Song Z, Gu FF, Guo XK, Ni YX, He P, Han LZ. 2017. Antimicrobial resistance and molecular characterization of *Staphylococcus aureus* causing childhood pneumonia in Shanghai. Front Microbiol 8:455. https://doi.org/10.3389/fmicb.2017.00455.

86. Yan X, Wang B, Tao X, Hu Q, Cui Z, Zhang J, Lin Y, You Y, Shi X, Grundmann H. 2012. Characterization of *Staphylococcus aureus* strains associated with food poisoning in Shenzhen, China. Appl Environ Microbiol 78:6637–6642. https://doi.org/10.1128/AEM.01165-12.

87. Vossenkuhl B, Brandt J, Fetsch A, Käsbohrer A, Kraushaar B, Alt K, Tenhagen B-A. 2014. Comparison of *spa* types, SCC*mec* types and antimicrobial resistance profiles of MRSA isolated from turkeys at farm, slaughter and from retail meat indicates transmission along the production chain. PLoS One 9:e96308. https://doi.org/10.1371/journal.pone.0096308.

88. Neela V, Mohd Zafrul A, Mariana NS, van Belkum A, Liew YK, Rad EG. 2009. Prevalence of ST9 methicillin-resistant *Staphylococcus aureus* among pigs and pig handlers in Malaysia. J Clin Microbiol 47:4138–4140. https://doi.org/10.1128/JCM.01363-09.

89. Baba K, Ishihara K, Ozawa M, Tamura Y, Asai T. 2010. Isolation of methicillin-resistant *Staphylococcus aureus* (MRSA) from swine in Japan. Int J Antimicrob Agents 36:352–354. https://doi.org/10.1016/j.ijantimicag.2010.06.040.

90. Larsen J, Imanishi M, Hinjoy S, Tharavichitkul P, Duangsong K, Davis MF, Nelson KE, Larsen AR, Skov RL. 2012. Methicillin-resistant *Staphylococcus aureus* ST9 in pigs in Thailand. PLoS One 7:e31245. https://doi.org/10.1371/journal.pone.0031245.

91. Chen CJ, Lauderdale TY, Lu CT, Chuang YY, Yang CC, Wu TS, Lee CY, Lu MC, Ko WC, Huang YC. 2018. Clinical and molecular features of MDR livestock-associated MRSA ST9 with staphylococcal cassette chromosome *mec*XII in humans. J Antimicrob Chemother 73:33–40. https://doi.org/10.1093/jac/dkx357.

92. Liu Q-Z, Wu Q, Zhang Y-B, Liu M-N, Hu F-P, Xu X-G, Zhu D-M, Ni Y-X. 2010. Prevalence of clinical methicillin-resistant *Staphylococcus aureus* (MRSA) with high-level mupirocin resistance in Shanghai and Wenzhou, China. Int J Antimicrob Agents 35:114–118. https://doi.org/10.1016/j.ijantimicag.2009.09.018.

93. Peng Q, Hou B, Zhou S, Huang Y, Hua D, Yao F, Shu Qian Y. 2010. Staphylococcal cassette chromosome mec (SCC*mec*) analysis and antimicrobial susceptibility profiles of methicillin-resistant *Staphylococcus aureus* (MRSA) isolates in a teaching hospital, Shantou, China. Afr J Microbiol Res 4:844–848.

94. Wannet W, Heck M, Pluister G, Spalburg E, Van Santen M, Huijsdans X, Tiemersma E, De Neeling A. 2004. Panton-Valentine leukocidin positive MRSA in 2003: the Dutch situation. Euro Surveill 9:3–4. https://doi.org/10.2807/esm.09.11.00484-en.

95. Yu F, Chen Z, Liu C, Zhang X, Lin X, Chi S, Zhou T, Chen Z, Chen X. 2008. Prevalence of *Staphylococcus aureus* carrying Panton–Valentine leukocidin genes among isolates from hospitalised patients in China. Clin Microbiol Infect 14:381–384. https://doi.org/10.1111/j.1469-0691.2007.01927.x.

96. Schaumburg F, Alabi AS, Peters G, Becker K. 2014. New epidemiology of *Staphylococcus aureus* infection in Africa. Clin Microbiol Infect 20:589–596. https://doi.org/10.1111/1469-0691.12690.

97. Glasner C, Pluister G, Westh H, Arends JP, Empel J, Giles E, Laurent F, Layer F, Marstein L, Matussek A, Mellmann A, Pérez-Vásquez M, Ungvári E, Yan X, Žemličková H, Grundmann H, van Dijl JM. 2015. *Staphylococcus aureus spa* type t437: identification of the most dominant community-associated clone from Asia across Europe. Clin Microbiol Infect 21:163.e1–163.e8. https://doi.org/10.1016/j.cmi.2014.09.010.

98. Kadariya J, Smith TC, Thapaliya D. 2014. *Staphylococcus aureus* and staphylococcal food-borne disease: an ongoing challenge in public health. Biomed Res Int 2014:827965. https://doi.org/10.1155/2014/827965.

99. He L, Zheng HX, Wang Y, Le KY, Liu Q, Shang J, Dai Y, Meng H, Wang X, Li T, Gao Q, Qin J, Lu H, Otto M, Li M. 2018. Detection and analysis of methicillin-resistant human-adapted sequence type 398 allows insight into community-associated methicillin-resistant *Staphylococcus aureus* evolution. Genome Med 10:5. https://doi.org/10.1186/s13073-018-0514-9.

100. Yang H, Paruch L, Chen X, van Eerde A, Skomedal H, Wang Y, Liu D, Liu Clarke J. 2019. Antibiotic application and resistance in swine production in China: current situation and future perspectives. Front Vet Sci 6.

101. Reynolds L, McKee M. 2009. Factors influencing antibiotic prescribing in China: an exploratory analysis. Health Policy 90:32–36. https://doi.org/10.1016/j.healthpol.2008.09.002.

102. Hyun JC, Kavvas ES, Monk JM, Palsson BO. 2020. Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. PLoS Comput Biol 16:e1007608. https://doi.org/10.1371/journal.pcbi.1007608.

103. Wheeler NE, Reuter S, Chewapreecha C, Lees JA, Blane B, Horner C, Enoch D, Brown NM, Estée Török M, Aanensen DM, Parkhill J, Peacock SJ. 4 September 2019. Contrasting approaches to genome-wide association studies impact the detection of resistance mechanisms in *Staphylococcus aureus*. bioRxiv https://doi.org/10.1101/758144.

104. Li X, Huang T, Xu K, Li C, Li Y. 2019. Molecular characteristics and virulence gene profiles of *Staphylococcus aureus* isolates in Hainan, China. BMC Infect Dis 19:873. https://doi.org/10.1186/s12879-019-4547-5.

105. Schroeder M, Brooks BD, Brooks AE. 2017. The complex relationship between virulence and antibiotic resistance. Genes 8:39. https://doi.org/10.3390/genes8010039.

106. Guillard T, Pons S, Roux D, Pier GB, Skurnik D. 2016. Antibiotic resistance and virulence: understanding the link and its consequences for prophylaxis and therapy. Bioessays 38:682–693. https://doi.org/10.1002/bies.201500180.

107. Roux D, Danilchanka O, Guillard T, Cattoir V, Aschard H, Fu Y, Angoulvant F, Messika J, Ricard J-D, Mekalanos JJ, Lory S, Pier GB, Skurnik D. 2015. Fitness cost of antibiotic susceptibility during bacterial infection. Sci Transl Med 7:297ra114. https://doi.org/10.1126/scitranslmed.aab1621.

108. CLSI. 2015. Performance standards for antimicrobial susceptibility testing. CLSI document M100-S25. Clinical and Laboratory Standards Institute, Wayne, PA.

109. Roch M, Gagetti P, Davis J, Ceriana P, Errecalde L, Corso A, Rosato AE. 2017. Daptomycin resistance in clinical MRSA strains is associated with a high biological fitness cost. Front Microbiol 8:2303. https://doi.org/10.3389/fmicb.2017.02303.

110. Li H. 2019. Readfq. https://github.com/lh3/readfq#readme.

111. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

112. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

113. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

114. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

115. Jolley K, Bray J, Maiden M. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.

116. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrico JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13:87. https://doi.org/10.1186/1471-2105-13-87.

117. Bartels MD, Petersen A, Worning P, Nielsen JB, Larner-Svensson H, Johansen HK, Andersen LP, Jarløv JO, Boye K, Larsen AR, Westh H. 2014. Comparing whole-genome sequencing with Sanger sequencing for *spa* typing of methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol 52:4305–4308. https://doi.org/10.1128/JCM.01979-14.

118. Kaya H, Hasman H, Larsen J, Stegger M, Johannesen TB, Allesoe RL, Lemvigh CK, Aarestrup FM, Lund O, Larsen AR. 2018. SCC*mec*Finder, a web-based tool for typing of staphylococcal cassette chromosome *mec* in *Staphylococcus aureus* using whole-genome sequence data. mSphere 3:e00612-17. https://doi.org/10.1128/mSphere.00612-17.

119. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

120. Côrtes MF, Botelho AM, Almeida LG, Souza RC, de Lima Cunha O, Nicolás MF, Vasconcelos AT, Figueiredo AM. 2018. Community-acquired methicillin-resistant *Staphylococcus aureus* from ST1 lineage harboring a new SCC*mec* IV subtype (SCC*mec* IVm) containing the *tetK* gene. Infect Drug Resist 11:2583–2592. https://doi.org/10.2147/IDR.S175079.

121. Rolo J, Worning P, Nielsen JB, Bowden R, Bouchami O, Damborg P, Guardabassi L, Perreten V, Tomasz A, Westh H, de Lencastre H, Miragaia M. 2017. Evolutionary origin of the staphylococcal cassette chromosome *mec* (SCC*mec*). Antimicrob Agents Chemother 61:e02302-16. https://doi.org/10.1128/AAC.02302-16.

122. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer, New York, NY.

123. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47:D687–D692. https://doi.org/10.1093/nar/gky1080.

124. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu S, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935.

125. Loughman A, Sweeney T, Keane FM, Pietrocola G, Speziale P, Foster TJ. 2008. Sequence diversity in the A domain of *Staphylococcus aureus* fibronectin-binding protein A. BMC Microbiol 8:74–74. https://doi.org/10.1186/1471-2180-8-74.

126. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

127. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

128. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.

129. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016.

130. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032.

131. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. https://doi.org/10.1186/1471-2148-7-214.

132. Kaplinski L, Lepamets M, Remm M. 2015. GenomeTester4: a toolkit for performing basic set operations - union, intersection and complement on k-mer lists. Gigascience 4:58. https://doi.org/10.1186/s13742-015-0097-y.

133. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953.

134. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. 2011. Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830.

135. Aun E, Brauer A, Kisand V, Tenson T, Remm M. 2018. A k-mer-based method for the identification of phenotype-associated genomic biomarkers and predicting phenotypes of sequenced bacteria. PLoS Comput Biol 14:e1006434. https://doi.org/10.1371/journal.pcbi.1006434.

136. Gerstein M, Sonnhammer EL, Chothia C. 1994. Volume changes in protein evolution. J Mol Biol 236:1067–1078. https://doi.org/10.1016/0022-2836(94)90012-4.

137. Méric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, Kittiwan N, Tadee P, Post V, Lamble S, Bowden R, Bray JE, Morgenstern M, Jolley KA, Maiden MCJ, Feil EJ, Didelot X, Miragaia M, de Lencastre H, Moriarty TF, Rohde H, Massey R, Mack D, Corander J, Sheppard SK. 2018. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. Nat Commun 9:5034. https://doi.org/10.1038/s41467-018-07368-7.