Epidemiology and Genomic Diversity of *Staphylococcus aureus* in Humans and Pigs in Kenya

VINCENT KIPLANGAT BETT

Fitzwilliam College, University of Cambridge

Wellcome Trust Sanger Institute

This dissertation is submitted for the degree of Master of Philosophy

August, 2018

ABSTRACT

Background

Staphylococcus aureus is an important pathogen of public health concern because of the emergence of methicillin resistant and multidrug resistant strains that can colonize and cause infections in humans and animals. However, there is little knowledge of pathogen characteristics and the circulating clones in low and middle-income countries where the burden of staphylococcal disease is often high. Here, whole genome sequencing data was used to determine the presence of shared clonal lineages, antibiotic resistance, virulence genes and the phylogenetic relatedness of 23 Kenyan strains. In addition, the correlation of certain resistance and virulence genes with phylogenetic lineages and the genetic relatedness of Kenyan isolates with respect to those of other countries were assessed.

Methods

Ninety-four isolates sampled from Kiambu county, Kenya, between October 2015 and August 2016, were randomly selected for sequencing using the llumina-B HiSeq X10 150 bp platform at the Wellcome Trust Sanger Institute. The 23 genomes (from 12 farmers and 11 pigs in 15 different homesteads) that passed minimum quality control thresholds, were used for analyses, in combination with 126 public genomes from the same lineages as defined by multi-locus sequence typing (MLST) results.

Results

The collection of *Staphylococcus aureus* isolates from Kiambu county was highly diverse represented by 9 Sequence Types (STs). The four major STs are ST188, ST789, ST25 and ST580, and were present in both the hosts. There were no genetic clonal lineage or specific classes of antimicrobial resistance genes that could be associated with either host. Notably, the presence of phages that carried human immune evasion gene clusters (IECs) in the majority of the strains suggests that the colonization by *S. aureus* could be of human origin. Comparison with public genomes revealed that the Panton Valentine leucocidin (lukF/S-PV) genes and enterotoxins genes clusters (*egc*) could be conserved in ST152 and ST25 lineages respectively. Furthermore, analysis of ST580 genomes with the double locus variant, livestock-associated ST398 lineage identified ST580 (2 pigs and 1 human) strains co-segregating with human associated clade of ST398 lineage based on the distribution of their accessory genes. The global population structure analysis also showed that the humans and

pig strains of this study are related to other African genomes, suggesting that Kenyan isolates could represent lineages circulating in the continent. Taken together, this study provides the first glimpse into the genomic diversity of *S. aureus* in Kenya and highlights the need for future genomic epidemiological surveillance using large datasets sampled from multiple hosts across many other collaborating countries.

ACKNOWLEDGEMENTS

It is with great pleasure and honor to take this special opportunity to thank my fantastic supervisor David Aanensen for his continual advice, encouragement and generous support throughout my Masters project. I feel privileged to be with his excellent team in learning and getting guidance in world of genomics. I extend my deepest gratitude to Silvia Argimon for her enthusiastic day-to-day guidance, attention to details and her commitment in assisting me to end this Master. I am also greatly indebted to Monica Abrudan and Sophia David for their kind guidance and generous time in transferring bioinformatics skills to me. I owe a great debt of gratitude to thesis committee members Mark A. Holmes, Sam Kariuki and Stephen Bentley for their expertise insights in genomics.

I am grateful to Wellcome Trust Sanger Institute Masters studentship, graduate office for giving an opportunity to realize not only my childhood dreams of studying abroad but also being in top globally excellent University. Special thanks to Annabel Smith, Christina Hedberg-Delouka and Sanger committee of graduate studies for their consistent advises, support and close monitoring of my progress.

I am greatly indebted to KEMRI staffs led by John Ndemi Kiiru for the generation of the study isolates and carrying out the DNA extraction. Many thanks to WTSI core sequencing department for library preparation and sequencing of the samples. I would like to express much appreciation to Pathogen Informatics team for their pipeline tools that were used in most analyses of this thesis. Many thanks to WTSI IT department for their services offered during the course of the study. I appreciate insights feedback of Matthew Holden and Ewan Harrison.

My deepest appreciation goes to Centre for Genomics Pathogen Surveillance team; Carol Churcher, Dawn Muddyman, Victoria Cohen, Corin Yeats, Ben Taylor, Simon Harris, Khalil AbuDahab and Richard Goater, who were very instrumental in my stay and learning at Sanger. I am very grateful to Fitzwilliam College for the accommodation and organizing social event activities which gave me good experiences about norms, culture and traditions of the University. Special thanks to my college tutor Jonathan Cullen and Fitzwilliam college chapel priest Revd Helen Arnold for their concern support during the terms.

Finally, I would like to offer my special thanks to my mother Gladys Chebet, family members, relatives and friends for their continual encouragement, prayers and support. And

special dedication of this thesis to my grandmother Grace Byomdo for her constant blessings and support during my childhood life.

DECLARATION

I declare that this thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

I declare this thesis is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. I further state that no substantial part of dissertation has already been submitted or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text

I declared that this thesis does not exceed 20,000 words prescribed in Special Regulation for MPhil in Biological Science.

Sign …….. Date ……..

PREFACE

The collection of study samples and DNA extraction were carried out by our collaborators at KEMRI led by Dr. John Ndemi Kiiru, KEMRI Kenya. WTSI core sequencing department completed sequencing of the DNA of the sampled strains. The analyses for the results of this dissertation were done by myself using mostly the pipeline developed by WTSI Pathogen Informatics pipeline as referenced in the texts. Public genomes that were included to extend the study were all downloaded by CGPS team except 6 strains which I retrieved myself from NCBI and 3 genomes that were donated by Dr. Dorota Jamrozy.

Table of Contents

LIST OF ABBREVIATIONS

ACCTRAN	Accelerated Transformation
AMR	Antimicrobial Resistance
AST	Antibiotic Susceptibility Tests
BAM	Binary Alignment Map
CC	Clonal Complex
CDS	Coding Sequences
DNA	Deoxyribonucleic Acid
G + C	Guanine-Cytosine
HIV	Human Immunodeficiency Virus
STs	Sequence Types
KEMRI	Kenya Medical Research Institute
Mb	Megabytes
MLST	Multi-locus Sequence Types
MSSA	Methicillin Susceptible *Staphylococcus aureus*
MRSA	Methicillin Resistant *Staphylococcus aureus*
NCBI	National Centre for Biotechnology
PVL	Panton-Valentine leucocidin
RAxML	Randomized Axelerated Maximum Likelihood
SAM	Sequence Alignment Map
SNPs	Single Nucleotides Polymorphisms
Spa	Protein A
VCF	Variant Call Format
WHO	World Health Organization
WGS	Whole Genome Sequencing
WTSI	Wellcome Trust Sanger Institute
GFF	General File Format

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

THESIS SUMMARY

This thesis is subdivided into five chapters. Chapter (1) highlights previous studies about biology of *Staphylococcus aureus* and the epidemiological findings while identifying knowledge gaps where this study could provide evidence for additional information. Chapter (2) illustrate the summary of methods used in analyses of the results. Chapter (3) describes results of quality control analyses, phylogenetic investigations and molecular characterization of Kenyan isolates. In addition, it outlines phylogenetic relationship of Kenyan strains with global public genomes. Chapter (4) highlights the discussion of the results in comparison with previous studies. Chapter (5) summarizes the key outcomes of the study and the future perspective of the study.

1.1 General Characteristics of the Genus Staphylococcus

The term 'Staphylococcus' has been in use since 1880s when Sir Alexander Ogston discovered that the post-operative acute abscesses among his patients were caused by grape-like clusters micro-organisms (Ogston, 1881). This term is now used as the genus name to define species that have a similar 16S RNA sequence, are facultative anaerobic and catalase positive (Humphreys, 2012). These species are Gram-positive cocci with a diameter of up to 2.0 μm and have low G + C content between 30-40%. In addition, they tolerate a high salt environment of up to 15 percent concentration and can survive in dry conditions (Todah, 2008). Some species are hemolytic, and some produce coagulase enzymes. The coagulase negative (CONS) species such as *Staphylococcus saprophiticus* are often saprophytes on the hosts but can sometimes cause opportunistic infections. *Staphylococcus aureus* (*S. aureus*) and *Staphylococcus epidermidis* are of clinical importance and are the best characterized of the more than 49 species and 26 subspecies under this genus as of 2018 (Kim et al., 2018).

1.2 Colonization, Infections and Diseases Caused by *S. aureus*

Previous studies have demonstrated that *S. aureus* is persistently carried in the noses of up to 20% of the healthy populations while 60% are recurrent carriers who do not develop symptoms (van Belkum et al., 2009, Peacock et al., 2001). Other anatomical sites colonized by *S. aureus* include the throat, uro-genital tract in women, intestines, respiratory tract, bones or heart valves (Acton et al., 2009, Kahl, 2010). *S. aureus* has been associated with many diseases such as benign skin infections in the community and life-threatening infective endocarditis, osteomyelitis, bacteraemia, pneumoniae and septicaemia in the hospitals (Klein et al., 2007).

S. aureus colonization can predispose the host for the development of diseases. This has been validated in previous studies undertaken in different settings such as among HIV positive patients (Weinke et al., 1992), surgical patients with wounds infections (Weinstein, 1959) and patients undergoing dialysis in the hospitals (Pignatari et al., 1990). However, elimination of *S. aureus* in nasal carriage can reduce incidence and prevalence of nosocomial infections (Peacock et al., 2001)

S. aureus can also colonize and infect farm animals and cause a significant impact on farmers' economy and animal welfare. For instance, economic loss associated with mastitis account for up to 40% of the cost of production in dairy cattle in the United Kingdom (UK) (Bradley, 2002, Holmes and Zadoks, 2011). In the pig industry, *S. aureus* cause exudative dermatitis (van Duijkeren et al., 2007).

1.3 Treatment and Antibiotic Resistance

The choice of antibiotic treatment strategies in combating *S. aureus* infections may be influenced by the type and site of infections, resistance profile of *S. aureus* in the region, availability and cost of the drug, as well as the host type (Siddiqui and Whitten, 2018).

Oral antibiotics such as trimethoprim, those of penicillin family, tetracyclines and linezolid are used for management of uncomplicated skin and soft tissues infections (SSTIs) and bactaeremia in humans (Ruhe and Menon, 2007, Nathwani et al., 2008). However, for complicated cases that can lead to hospitalization or that occur among hospitalized patients, the drugs of choice are normally intravenous vancomycin or daptomycin or flucloxacillin monotherapy for cases caused by methicillin susceptible *S. aureus* (MSSA) (Sutherland et

al., 1970). Skin abscesses infections can be treated by the adequate surgical drainage (Rajendran et al., 2007).

The phenotypic resistance of bacteria to antibiotic therapy poses a significant challenge to the treatment of infections in human and agricultural medicine worldwide, resulting in a limited number of treatments available for severe infections. Nearly 25,000 deaths occur annually in developed countries such as in Europe and the United States as a result of the high burden of antimicrobial resistance (AMR) according to the World Health Organization global surveillance report of antibiotic resistance. This estimate could be higher in developing countries where the health systems are weak (http://www.who.int/drugresistance/documents/surveillancereport/en/). Of interest, nearly a half of these deaths were caused by *S. aureus* infections. This could be because some *S. aureus* strains quickly respond to the introduction of new antibiotics (Noble et al., 1992).

AMR in *S. aureus* may occur due to inactivation or protection of the antibiotic target sites, destruction or reduction in the intracellular amount of antibiotic and alteration of their membrane permeability (Blair et al., 2015). Some *S. aureus* strains may overexpress efflux pumps so as to remove harmful substances including drugs from the cells (Floyd et al., 2010).

AMR in *S. aureus* was first reported for penicillin, shortly after its introduction to clinical use in the 1940s (Rammelkamp and Maxon, 1942, Plough, 1945). The *S. aureus* may have acquired an extra-chromosomal plasmid which produced β-lactamase enzymes that directly inactivate the antibiotic through hydrolysis of the amide ring of the β-lactam ring (Novick and Bouanchaud, 1971). This resistance led to the discovery of methicillin in 1959 as an alternative treatment for penicillin-resistant *S. aureus*, but methicillin resistant (MRSA) was shortly identified in three *S. aureus* strains from among over 5000 clinical isolates tested in UK (Jevons, 1961). The *mec*A gene that encodes for the penicillin binding protein PBP2a was primarily integrated into staphylococcal cassette chromosome (SCCmec), thus allowing *S. aureus* to continue synthesizing the peptidoglycan cell wall despite inhibition of the normal cell PBP2a by the β-lactam antibiotic (Katayama et al., 2000). In 2011, *mec*C which shared almost 70% nucleotide sequence to *mec*A was discovered to be also causing methicillin resistance (Garcia-Alvarez et al., 2011). The continued rise in detection of the *mec*C gene in *S. aureus* strains poses a significant challenge in developing countries that rely

on conventional microbiology assays for diagnosis because *mec*C cannot be detected by the normal PCR or latex agglutination assays used for current routine testing of MRSA (Stegger et al., 2012). Both *mec*A and *mec*C are carried on a large SCC*mec* element. In addition to *mec* genes, SCC*mec* elements also harbor *ccr* genes, regulatory genes, and accessory genes. There are eleven different types of SCC*mec* elements that had been identified up to date (Svensson et al., 2011). The assignment of SCC*mec* elements into types depend on the combination of *mec* and *ccr* complex genes. Further subdivision of SCCmec types into subtypes depend on variations within joining regions (J-regions) which are classified into subgroups, J1-3 (Svensson et al., 2011).

Another important mechanism of development of AMR is mutational changes to the protein target sites such as DNA gyrase A and topoisomerase in the case of fluoroquinolone resistance or multiple bases mutations in the 23S ribosomal subunit drug binding sites for the linezolid resistant *S. aureus* (Tsiodras et al., 2001, Meka et al., 2004).

Of particular concern in AMR is the continued rise in emergence of vancomycin intermediate *S. aureus* (VISA) and heterogenous VISA (hVISA since it causes staphylococcal infections where use of vancomycin as last resort for treatment fails (Claeys et al., 2016, Zhang et al., 2015). In addition, its prevalence is under-reported because of lack standard detection techniques (Zhang et al., 2015). The *vanA* operon genes which confer resistance to vancomycin are normally carried in the plasmid copies of the transposon Tn1546 which has been suggested to have been acquired from vancomycin resistant *Enterococcus faecalis* (Chang et al., 2003).

Antimicrobials have a wide range of use in veterinary medicine ranging from treatments of infections, promotion of growth, prophylaxis, to prevention of spread of infection in case of an outbreaks (Woolhouse et al., 2015). Notably, there is no clear distinction of antibiotic use between human and veterinary medicine. According to the World Organization for Animal Health (OIE) international committee, some specific classes of antimicrobials that are of great importance in veterinary medicine are also used in human medicine for treatments (OIE, 2007). The resistance genes such as tetracycline genes *tet*(*M*), lincosamides (*lnuA*), macrolides (*ermC*) associated with livestock associated MRSA CC398 lineage are also commonly identified in other lineages of *S. aureus* of human origin (Kadlec et al., 2012).

The increased prevalence and significance of multidrug resistant (MDR) *S. aureus* worldwide and their spread even outside the healthcare settings necessitate the need to undertake research studies with the aim of understanding the source and basis of resistance, prevalence of antibiotic resistance and to inform the healthcare profession in their therapeutic management of the patients.

1.4 *S. aureus* as a Significant Public Health Problem in Africa

Staphylococcal disease is a major public health concern in both developed and developing countries. *S. aureus* is listed among the highest priority bacterial pathogens that cause high morbidity and mortality worldwide according to recent publication of World Health Organization (W.H.O) on use of antibiotics (Tacconelli et al., 2018). *S. aureus* in well-resourced environments have been extensively analyzed with respect to source, type of infection and treatment strategies (Wertheim et al., 2004). In addition, the majority of these countries screen patients for MRSA carriage on admission to hospitals (Wertheim et al., 2004). Furthermore, in developed countries such as Netherlands where MRSA prevalence are low (< 3%), have 'search and destroy policies' (Vos et al., 2005) in addition to carrying out longitudinal studies in a community setting for epidemiological surveillance purposes (Bergstrom et al., 2013). This contrasts sharply with developing countries, where it is considered an insignificant cause of morbidity and mortality and yet the number of deaths due to staphylococcal disease is considerably higher than in developed countries (Nickerson et al., 2009). The emphasis in infectious diseases is on malaria, HIV and tuberculosis instead (Herrmann et al., 2013). Worryingly, high burden among the patients are co-infections of these diseases with Gram positive cocci including *S. aureus* (Herrmann et al., 2013). For instance, there is high prevalence of *S. aureus* among HIV positive paediatric patients (Lemma et al., 2015, Berkley et al., 2009) and MRSA carriage among hospitalized tuberculosis patients in hospitals in HIV endemic area (Heysell et al., 2011).

Importantly, MRSA prevalence appears to have been increasing in Africa since 2000 (Falagas et al., 2013). Moreover, Sub-Saharan African countries have different distribution of clonal lineages of *S. aureus* (Schaumburg et al., 2014a). This could possibly be ascribed to a number of factors such as varied socio-economic status, cultural and climatic conditions (Lozano et al., 2016b). Furthermore, poor infection control and patient treatment

management, crowded living conditions, unhygienic conditions, high temperatures and humidity (Wang et al., 2013), HIV infection (Kinabo et al., 2013), and close contacts of animals with humans (Schaumburg et al., 2012) may facilitate transmissions of *S. aureus.* Hence, efforts are desperately needed for epidemiology and population structure studies in these resource-limited environments.

Recent studies in Sub-Saharan African countries have mostly been limited to investigating prevalence and colonization levels of *S. aureus* in infections and morbidity. They relied on conventional microbiological and clinical interpretation performed on traditional methods such as culture sensitivity and low discriminatory tools including Antibiogram (specific profile of antibiotic panels). As a consequence, there is little data on emerging and circulating clones, evolutionary relationships and origins of major clonal lineages and their molecular biology.

1.5 One Health Concept

The increased use of antibiotics as part of animal feeds to promote growth, and emergence and re-emergence of zoonotic infectious diseases including *S. aureus* infection remains a serious concern for humankind. Zoonotic pathogens cause approximately 61% of human infections and 75% of emerging diseases in humans (Taylor et al., 2001). Intriguingly, the majority of emerging zoonotic infectious diseases in humans could be traced to animal origin (Woolhouse et al., 2005). The 'host-switch' of these zoonotic and anthroponotic pathogens is accompanied by spill-over in their biological and genetic factors including antibiotic resistance and virulence factors determinants (Shepheard et al., 2013). These could aid in invading a new host and promote emergence of a new clone when exposed to a different environment as a result of deletions, insertions, recombination and acquisition of mobile genetic elements (Ben Zakour et al., 2008).

The extensive consumption of antibiotics in agricultural medicine, even in developed countries where antibiotic use is highly regulated, make it necessary to understand the impact on human healthcare systems, and to design policies that give clear guidelines regarding antibiotic usage in veterinary (Grave et al., 2010). For this reason, the WHO proposed a One Health Initiative to bring together collaborators working in human, environment and animal sectors with the main objective of understanding the emergence, spread and prevention of

antibiotic resistance of zoonotic pathogens including *S. aureus* in animals and the potential impact on human health.

Furthermore, the United Nations Food and Agriculture Organization (FAO) has estimated a worldwide rise in demand for livestock meat from approximately 230 million metric tons in 2000 to 300 million in 2020 in which the majority would come from developing countries (http://hdl.handle.net/10947/1622). Consequently, there will likely be an increase in zoonotic emerging infectious diseases, due to high plethora of factors which include an increased population density, close contacts with animals in the farms, coupled with inadequate technical expertise and infrastructure to combat disease outbreaks (Declercq et al., 2008). On top of these, the majority of the population are deficient in immune system (Gebreyes et al., 2014). Therefore, there is an urgent need to address this crucial issue through research on the sharing of clonal lineages, adaptations and evolutionary relationships of *S. aureus* found amongst humans and animals.

1.6 Transmission of *S. aureus* Between Humans and Livestock

Inter-species transmission of *S. aureus* between humans, companion animals and livestock have been documented previously especially in developed countries. The first detection of MRSA in animals was in milk from Belgian cows with mastitis in 1972 (Devriese et al., 1972). Since then occurrence of MRSA has been reported from a wide variety of animals including chickens (Sallam et al., 2015), cats and dogs (Loeffler et al., 2005), and swine (Chuang and Huang, 2015).

Livestock associated *S. aureus* clonal complexes CC398 which is frequently isolated from pigs was identified for the first time in the Netherlands in 2004 (Voss et al., 2005) and several countries, especially in Europe and North America, have reported it afterwards (Armand-Lefevre et al., 2005, Cuny et al., 2015, Mediavilla et al., 2012). Of note, infections following carriage of CC398 among farmers and close family members are rising (Schijffelen et al., 2010). Interestingly, ST398 strains with livestock associated characteristics are not transmitted easily between humans but these studies demonstrated that the colonization levels reduced considerably among the farmers upon withdrawal of the strains' reservoirs (Graveland et al., 2011, van Cleef et al., 2011). This suggests that the pigs act as an important

reservoir for development of virulence factors and antimicrobial resistance determinants in the *S. aureus* for subsequent infections in humans (Fitzgerald, 2012, Price et al., 2012a).

Voss et al. demonstrated that there is higher chance (760 times) for MRSA to colonize pig farmers than the Dutch general populations among the patients admitted to hospitals in the Netherlands (Voss et al., 2005). A similar previous study conducted in France provided further evidence to support transmission in which they found that the farmers and pigs were colonized by the same sequence type (ST) ST9, ST398 and ST433, but, interestingly, these were not seen in control individuals (Armand-Lefevre et al., 2005). In addition, Rinsky et al demonstrated that the workers at industrial livestock operations in the USA carried nasal *S. aureus* with similar characteristics of livestock-associated lineages such as tetracycline resistance and being *scn*-negative, which was not observed among workers with no contact with livestock (Rinsky et al., 2013). This high risk might be because close contact could facilitate cross inter-species transmission especially when animals are infected with heavy loads of antimicrobial agents, and also the ability of *S. aureus* to survive in the harsh and unfriendly environment (Le Loir et al., 2003).

Various studies which have been undertaken in pigs suggest that different livestock-associated clonal lineages circulate in different regions. For instance, ST398 is a major clone in Europe and North America while ST9 is frequently isolated in Asian countries (Chuang and Huang, 2015). In Africa, the population structure circulating in livestock is unclear. CC398 is yet to be reported in pigs, although ST398 has been detected in chicken samples in Tunisia (Chairat et al., 2015). Only two countries have demonstrated MRSA colonization in pigs, including Senegal and South Africa, in which prevalence rates were estimated to be 1.3% and 12.5%, respectively (Fall et al., 2012, Adegoke and Okoh, 2014). These were unexpectedly low compared to the 3- 80% LA-MRSA reported prevalence in developed countries.

Furthermore, there are limited epidemiological studies of *S. aureus* presence, distribution and molecular typing in animals available in Sub-Saharan Africa (Lozano et al., 2016b), presumably not because of low prevalence, but probably due to lack of technical expertise and laboratory facilities. In addition to that, genetic relatedness of human and pig *S. aureus* clones using high discriminatory power such as whole genome sequencing has not been investigated.

1.7 Conventional Molecular Typing Tools

Molecular typing schemes are of great relevance in epidemiological investigations of the spread of clones, the genetic variation of strains and the evolution of bacteria. These methods have evolved over the last years from phenotypic characterization, to typing techniques that utilize the most variable loci in the *S. aureus* genomes and, more recently, the whole genome. For instance, pulsed field gel electrophoreses (PFGE) was popular in 1980s and has been harnessed mainly in nosocomial outbreaks and patient-to-patient transmissions in a local clinical settings (Bannerman et al., 1995). It is based on the separation of a restriction digest of genomic DNA on a gel upon periodic application of electric current (Herschleb et al., 2007). However, it relies on reproducibility of gel patterns and is subsequently prone to error in interpretation of closely related clones with highly similar DNA band sizes and patterns. In addition, it is difficult for comparison between the laboratories, it is labour-intensive and time-consuming (Cookson et al., 1996, Murchan et al., 2003).

In 1998, a nucleotide sequence-based approach technique, multi-locus sequencing typing (MLST) that has merit over PFGE was proposed to offer relatively higher accuracy and improved portability of information between laboratories. MLST has been used to study the population structure of *S. aureus* and its international spread, and used in population studies of association of virulence and antibiotic resistance genes with certain lineages (Urwin and Maiden, 2003). The DNA fragments are submitted to the MLST website (www.mlst.net) and compared to the allelic profile of the ~450bp DNA sequences of seven housekeeping genes (*arcC, aroE, glpF, gmk, pta, tpi*, and *yqiL*) to define the sequence types (STs) (Enright et al., 2000). Point mutations within the seven sequence genes of the isolates in the same lineages give different, but closely related STs. Such isolates that share at least five of the seven sequence loci are grouped together into their respective Clonal Complexes (CCs) using a web-based algorithm called eBURST (Based Upon Related Sequence Types) (Feil et al., 2004). This is useful for grouping lineages that share a recent common ancestor and have similar genetic composition. Although it is useful in outbreaks investigations, MLST have relatively low resolution to differentiate recent evolution of bacterial clones of closely related strains that lack mutations at those seven housekeeping genes.

Another popular typing method in epidemiological analyses is *spa* sequence typing that has relatively good discriminatory power, ease of use and interpretation, and portable in sharing

of information between laboratories by submitting sequence data to the Ridom SpaServer (http://spaserver.ridom.de). It is highly convenient for rapid outbreak investigations in a hospital setting (Hallin et al., 2009). It relies on DNA sequencing of the polymorphic X that has 24-bp repeats and are found upstream of the C- terminal cell wall which is attached to the protein A gene (*spa*) (Guss et al., 1984). The sequence alignment variations used in assigning different *spa* type code arise due to variation in the number of repeats units and point mutations, duplication or deletions of the repeat units (Brigido Mde et al., 1991).

1.8 Whole Genome Sequencing Technologies and Application

The first publication of complete sequencing of a bacteriophage using the specific chain terminating inhibitors of DNA polymerase by Sanger et al. in 1977 has led to development of three generation sequencing technologies (Sanger et al., 1992). The landmark of sequencing was the introductions of automations of capillary sequencing of the first generation. Capillary sequencing separates specially end-labeled DNA fragments based on their sizes which are generated by sequencing by synthesis or degradation. It was used in the initial sequencing of the human genome, but it has drawbacks of being expensive, slow and labour-intensive (Schadt et al., 2010). This was the reason why the human genome project costed almost ten billion US dollars and took nearly ten years for it to be completed.

The lengthy period and high cost of sequencing involved in the completion of human genome project led to development of next generation sequencing (NGS). The commercially available NGS include Illumina, 454 Roche genome sequencer, SOLiD Life Technologies, Helico Biosciences, and Ion Torrent (Metzker, 2010). Multiple DNA fragments are sequenced in parallel through successive washing and scanning processes. During sample library preparation, index tagged adapters are linked reverse and forward to the isolates and this allows a large number of bacterial isolates to be sequenced on a single lane (Loman et al., 2012). This current project relies entirely on Illumina WGS analyses for the molecular characterization and inferring the transmission of *S. aureus* isolates between humans and pigs.

Traditional typing schemes such as MLST are based on variation in one or more genes that have a sequence size between 400 and 4000 nucleotides to assign strains to different lineages, while with WGS we can detect variation in the entire genome (2.8 Mbs for *S.*

aureus) (Price et al., 2013). WGS has proved to offer high resolution of closely related strains to a single base pair difference (Price et al., 2013).

In 2010, (Harris et al., 2010a) demonstrated for the first time use of whole genome sequencing (WGS) in bacterial genomics on a large dataset of a single MRSA lineage of ST239 to infer transmission over time, across continents, and within a certain hospital setting in Thailand. They found evidence of transmission among five patients (<14 SNPs) within the same hospital and also transmission of *S. aureus* across four continents. This suggested need for global surveillance. In addition, they estimated the evolutionary rate in the core genome of this MRSA lineage to be approximately 6 SNPs per genome per year and this could be used to infer the time to the most recent common ancestor between the two closely isolates in endemic and outbreak investigations (Harris et al., 2010a). They also identified that the point mutations of *spa* typing used to assign lineages to infer transmission in outbreaks investigations may not be as reliable as using a WGS-based phylogeny. For example, some of the isolates that were of the same *spa* type were genetically distant by WGS while others of different *spa* types were closely related.

The current application of WGS in bacterial genomics has shown that valuable results can be obtained through *in silico* prediction of phenotypic characteristics of antibiotic resistance. Aanensen et al (2016) demonstrated that the genotypic prediction of AMR may as well be as reliable as phenotypic antibiotic susceptibility testing for *S. aureus*. They found high concordance results (98%) between genotypic and gold standard phenotypic methods for 19 antibiotics tested against over 300 *S. aureus* isolates (Aanensen et al., 2016b). These small discrepancies could be as a result of loss of phage, plasmid or transposon carrying the resistance during and after sequencing and probably the choice of culture medium. This shows that i*n silico* prediction of resistance genes of genomic data offers the prospect of replacing the gold standard methods which are relatively laborious and expensive. This may only be realized if there is greater understanding of the association of presence of genes in the strains and their phenotypic expression.

Furthermore, WGS has also been used to validate the previous assumptions about the origin of certain *S. aureus* clones or genes. For example, Price et al (Price et al., 2012b) analyzed 89 sequenced isolates of ST398 from different continents and demonstrated that the livestock-associated MRSA ST398 may have originated from MSSA in humans, and

subsequently colonized and underwent adaptations in animals through loss of MGEs such as phages associated with immune evasion clusters and gain of tetracycline resistance. This supported previous studies that had identified the infections of humans with ST398 *S. aureus* isolates without having necessarily been in close contacts with livestock, and which lacked characteristics of livestock-associated lineages. WGS was also used to confirm previous findings that the resistance genes of the vancomycin-resistant *S. aureus* isolates may have been originated from enterococci through acquisition of a transposon (Kos et al., 2012).

Despite a lot of advantages of WGS and its prospect for its future use, it has drawbacks. For instance, the presence of contaminants in DNA isolates due to insufficient purifications in the selective media, or poor containment procedure in DNA preparation and extraction influence the quality of the genomic data results. In addition, it generates multiple short reads that need to be either mapped to the reliable reference or de novo assembled (Linderholm, 2016) each of which comes with limitations. Mapping-based assembly fails to make use of sequence regions that are not found in the reference genome while de novo assembly methods usually result in fragmented assemblies due to the inability to assemble repetitive regions. Furthermore, WGS data requires a lot of bioinformatics expertise in the analyses and interpretation of the results. The bioinformatics knowledge is at its nascent stage in developing countries and this together with the high cost of WGS, could be the primary reason previous studies on zoonotic transmission using WGS is skewed towards the developed countries.

There are efforts to reduce the cost of sequencing, develop ease-to-use bioinformatics tools and incorporate these tools into automated pipelines allowing use in hospital settings to guide routinely diagnosis and treatment strategies. Third generation sequencing technology (TGS) seems to offer better advantages over other generations for clinical application. They apply the same sequencing principle as NGS although the sequencing reactions is completed in only a matter of hours, relatively inexpensive, and yields longer reads (Loman et al., 2012). However, the current TGS has limitations and there is a lot of advancement required to improve throughput for it to be applied in large scale genomics studies and to lower the sequencing error rates.

1.9 *S. aureus* Genome and Population Structure

Completion of the first genome sequences in 2001 of two clinical isolates MRSA N315 and vancomycin-resistant Mu50 (Kuroda et al., 2001a), and publications of many more annotated genome sequences of *S. aureus* isolated from diverse settings (Holden et al., 2004b, Baba et al., 2002), combined with the decreasing costs of Illumina sequencing, has led to generation of thousands of publicly available genomes of *S. aureus*. Notably, there are more than forty thousand publicly *S. aureus* genomes that have been deposited in the European Nucleotide Archives (ENA) as of November 2017 (Petit Iii and Read, 2018).

The complete genome size of *S. aureus* is roughly 2.8 Mb in size and comprises of core genes that make up to 75% and 10 - 25% of accessory genes (Lindsay and Holden, 2004). *S. aureus* has nearly 2500 protein open reading frames and 32.0% G + C content (Holden et al., 2004a). The conserved genes mainly make up the core genome and encode for functional proteins that are fundamental for growth such as those that take part in metabolism, synthesis and replication of genetic materials, although quite a number of these genes still have unknown functions (Lindsay, 2014). The core genes may be present in more than 95% of all *S. aureus* species (Kuroda et al., 2001b). The point mutations and selection pressures of the conserved genes seems to be driving the evolution and emergence of new *S. aureus* lineages (Lindsay, 2014).

Accessory genomes mainly harbor mobile genetic elements (MGEs) that include Staphylococcal chromosomal cassettes (SCC), bacteriophages, plasmids, prophages pathogenicity islands and transposons. These MGEs contain virulence and antimicrobial resistance genes that could be vital for the colonization, survival, and fitness of the strains in the hosts (Lindsay, 2010). This accessory genome may have relatively higher G + C content than the core genome due to acquisition of genetic materials from other bacterial species, or the environment as a result of selective pressures to adapt to a new ecological niche (Ben Zakour et al., 2008).

Previous studies have validated that the exchange of MGEs between individual *S. aureus* strains is facilitated by the horizontal gene transfers that occur either through conjugation or transduction (Chambers and Deleo, 2009, Lindsay, 2010). Different MGEs seem to carry certain virulence and antibiotic genes which are distributed according to certain CCs and the hosts (Lindsay, 2010).

The *S. aureus* population structure is highly clonal without undergoing much extensive recombination, has worldwide distribution and is stable over time (Monecke et al., 2011). There are only a few dominant lineages of about ten that are known to cause diseases in humans. These clonal complexes include CC1, CC12, CC8, CC5, CC22, CC25, CC30, CC45, CC51 and CC15 (Feil et al., 2003). Only a few lineages seem to be dominant in animals such as CC151, CC97, CC126, CC133, CC771 in bovine hosts. Some of the lineages are shared across the diverse hosts and include CC8, CC22, CC25 and CC398 (Holmes and Zadoks, 2011). The success of these few dominant clones in different continents could be attributed to their ability to withstand pressures in the exposed environment. This is because they possess surface proteins and their regulators that could interact easily with the host, have insertion or variation regions within the core gene as well as presence or absence of certain genes and their ability to acquire SNPs within the core genes (Lindsay, 2010, Lindsay and Holden, 2004).

1.10 Virulence Factors of *S. aureus*

S. aureus has gained worldwide relevance in both human and animal health not only because of its antibiotic resistance but also because of its plethora of virulence factors that enables it to effectively colonize and cause infections (Zecconi and Scali, 2013). *S. aureus* produces two major categories of virulence protein factors ; surface proteins and exotoxins which are located either in the core or accessory genomes (Ballhausen et al., 2017). In addition, it has capsular polysaccharide that is linked to peptidoglycan, and acts to prevent phagocytic killing (Thakker et al., 1998) and as protective layer to withstand pressure from harsh environment.

1.11.1 Adherence factors (Surface Proteins)

The first stage of colonization of *S. aureus* is the adherence to the host cells. The cell wall anchor surface proteins initiate the first step by covalently attaching to the extracellular matrix which may contain collagens, fibronectin, and/or fibrinogen. This is a critical step in the successful colonization and survival of *S. aureus* to be a commensal organism in the host (Foster et al., 2014). However, there are other surface proteins that have other annotations apart from adherence and these have been grouped together under a term called microbial

surface component recognizing adhesive matrix molecules (MSCRAMMs) (J M Patti et al., 1994). Examples of these surface proteins include elastin binding protein (*ebp*), fibronectin-binding A (*FnBpA*), fibronectin-binding B (*FnBpB*), collagen-binding protein (*Cna*), clumping factor A (*clfA*) and clumping factor B (*clfB*) proteins (Lowy, 1998, Foster and Hook, 1998).

1.11.2 Exotoxins (Extracellular Enzymes)

The exotoxins which are expressed by most *S. aureus*, are essential for the survival and spread of these strains within the host. These exotoxins include deoxyribonucleases, proteases, lipases, hyaluronidase and collagenase (Dinges et al., 2000). Some exotoxins are secreted by specific strains to aid in their invasion, pathogenicity and suppression of immune systems of the hosts. These include hemolysins, the exfoliative toxins (*eta*, *etb*, *etd*), leucocidin, Panton-Valentine leucocidin (PVL) and pyrogenic toxin superantigens (PTSAgs) such as toxic shock syndrome toxin-1 (TSST-1) and staphylococcal enterotoxins (SEs) (Lina et al., 2004). In humans, PTSAgs cause the mitogenic impact on T cells leading to toxic shock syndrome resulting in symptoms such as rash, high fever, hypotension and multiple organ dysfunction (Novick et al., 2001).

The expression of both exotoxin and surface protein virulence factors are controlled by the accessory gene regulator (*agr*) (Yarwood and Schlievert, 2003). Therefore, *agr* of *S. aureus* plays a critical role in pathogenesis and infections of the hosts. Up to date, there are four types of *agr* groups (agrI to agrIV) which are determined based on agrD and agrC polymorphisms (Ostojic and Hukic, 2015).

The distribution of virulence factors seems to vary between isolates from different hosts, from asymptomatic carriers and infections, and between different geographic regions (Ostojic and Hukic, 2015, Ikawaty et al., 2010, Kadlec et al., 2009).

1.11.a The Staphylococcal Enterotoxins (SEs)

Staphylococcal food poisoning in humans has been a major public health concern worldwide. It occurs mainly as result of ingestion of food contaminated with *S. aureus* that produces enterotoxins (Hennekinne et al., 2012). These SEs are categorized into two; classical (SEA-SEE) and novel (SEG-SEY) enterotoxins groups (Argudín et al., 2010). Both have superantigenic activity to cause disease characterized by symptoms such as nausea,

abdominal cramps and diarrhoea (Pinchuk et al., 2010). Intriguingly, only small amounts of SEs are enough to be toxic in humans. Of note, they tolerate unfavorable environments such as exposure to low pH, high temperatures and chemical denaturation (Regenthal et al., 2017). Most of these SEs are harboured in the mobile genetic elements (MGEs) such as *sea* in temperate phages, *seb* in pathogenicity islands, *sec* in plasmid and *seg*, *sei*, *sem* in genomic islands (Fisher et al., 2018). The exchange of these MGEs between *S. aureus* could facilitate the spread, the ability to cause disease and the evolution of these strains (Argudín et al., 2010).

1.11.b Immune Evasion Cluster (IEC) Genes

Some *S. aureus* strains have phages that contain genes that encode for modulatory proteins of the human innate immune response. These genes include the staphylococcal complement inhibitor (*scn*), enterotoxin (*sea* and *sep*), staphylokinase (*sak*) and chemotaxis Inhibitor Protein (*chp*) and are integrated into β-hemolysin (*hlb*) converting bacteriophages (Christiane et al., 2006). These immune modulatory proteins are highly specific to humans (Koymans et al., 2017) and act synergistically together to prevent innate immune systems from destroying them thus aiding them in colonizing the host (Xia and Wolz, 2014).
Previous studies have demonstrated that transfer of *S. aureus* from humans to animals and vice versa is accompanied by the gain or loss of these virulence traits carried in the bacteriophages. This mechanism enables the pathogen to adapt to the host (Christiane et al., 2006). Notably, more than 90% of *S. aureus* isolated in humans from clinical settings are positive for these phages (Pantucek et al., 2004, van Wamel et al., 2006a). In contrast, these phages are nearly absent among animal isolates and this is demonstrated by high prevalence (66-92%) of Hlb positive strains among animal isolates (Verkaik et al., 2011b).

1.11.c Panton-Valentine Leukocidin

Panton-Valentine leucocidin (PVL) toxin is a bi-component, pore forming exoprotein, encoded by lukF-PV, and lukS-PV sub-unit genes. They are located on the temperate bacteriophages that are carried especially by CA-MRSA strains (Boyle-Vavra and Daum, 2006). PVL induce lysis of monocytes and neutrophil granulocytes. Recent studies have associated PVL with severe skin and soft tissue infections (Lina et al., 1999) and other life-

threatening diseases such as necrotizing hemorrhagic pneumonia especially among children and immune compromised patients (Gillet et al., 2002).

Epidemiological studies in Sub-Saharan Africa have reported high prevalence (17-74%) of PVL positive *S. aureus* among clinical isolates with majority carried by MSSA (Schaumburg et al., 2014a, Abdulgader et al., 2015). Despite PVL being specific to human neutrophils, it was identified in equal frequency in both chimpanzee and human isolates in Gabon suggesting that animals may act as a reservoir of this virulence factor in Africa (Abdel-moein et al., 2012).

1.12 OBJECTIVES OF THE STUDY

This chapter has highlighted that *S. aureus* is capable of colonizing variety of hosts and subsequently cause life-threatening infections. Notably, they easily developed resistance to any newly introduced antibiotic. Advancing one health concept through investigation of *S. aureus* between humans and pigs could help reduce the knowledge gaps of (1) clonal lineages of *S. aureus* circulating in humans and pigs in Kiambu county, Kenya, (2) antimicrobial resistance genes carried by *S. aureus* in humans and pigs, and importantly (3) identifying the virulence factors that could be associated to hosts and certain lineages. Even though previous studies in Sub-Saharan Africa have investigated transmission of *S. aureus* between humans and animals but most have used low discriminatory power typing methods which we believe use of whole genome sequencing could fill the knowledge gaps in terms of (3) inferring genetic relatedness of strains between the hosts and across the homesteads. Additionally, (4) understands the genetic relationship of Kenyan isolates with global strains when combine with publicly available genomes.

CHAPTER TWO

METHODS

2.1 Ethical Considerations

Ethical approval was obtained from Kenya Medical Research Institute (KEMRI) and the Department of Livestock and Fisheries of Kenya government to conduct sampling and collect metadata information from participating farmers and their corresponding pigs in Kenya, and later for the shipment of the DNA and implementation of the study in the United Kingdom.

2.2 Study Design and *S. aureus* Isolation

Between October 2015 to August 2016, 248 *Staphylococcus aureus* isolates were collected from the nares of farmers and their corresponding pigs from different homesteads spread throughout 9 districts of Kiambu County, Kenya. Kiambu has a population of 1.5 million persons in a 2500 km^2 and is adjacent to north of Nairobi, capital city of Kenya. From each homestead, up to a maximum of 3 farmers and 3 pigs from herd size that range between 5 and 125 were randomly selected for sampling. About 55% of the homesteads sampled were keeping pigs for commercial purpose.

Written informed consent was obtained from each participating person. Swine isolates were excluded from the study if the handler or the owner refused to be sampled or to provide information such as age, history of the usage of antimicrobial agents, number of pigs in a farm, presence of pets and other livestock, and size of the farm. Sampling was done by inserting sterile cotton swabs moistened with 0.85% sodium chloride into the nose of the host.

The samples were sent within 24 hours to the Kenya Medical Research Institute/Center for Microbiology Research for processing. The nasal swabs were enriched in a 5 ml Tryptone soy broth and incubated at 37^0C for 24 hours. On the following day, they were transferred to a differential media where they were plated on Mannitol Salt Agar (MSA) and incubated overnight at 37^0C for 18-24 hours. The presumptive positive yellow colonies were subjected

to Gram staining, and catalase (Oxoid, UK), oxidase (Oxoid, UK) and latex Staphaurex agglutination (Oxoid, UK) tests in order to confirm presence of *S. aureus* species.

Ninety-five isolates collected from 23 homesteads were randomly selected for DNA extraction, 47 from humans and 48 from pigs from 29 homesteads. All except 8 strains were paired by homesteads.

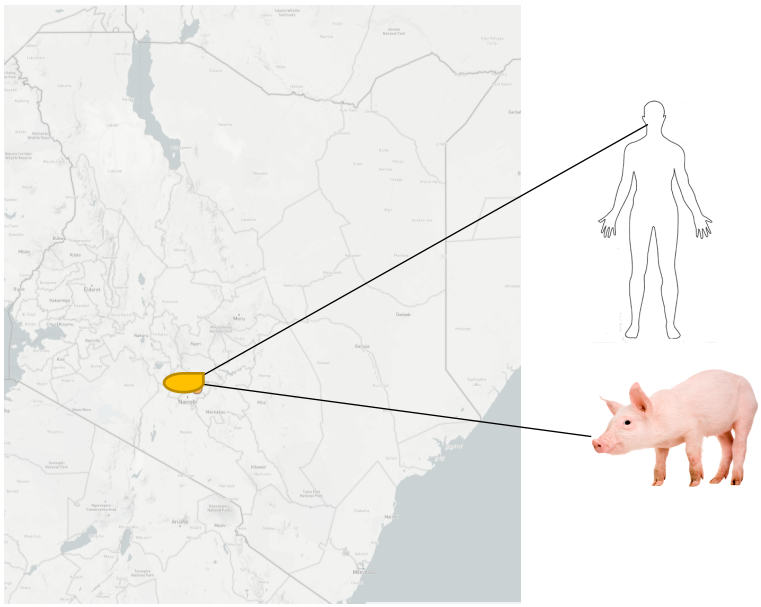


Fig. 1: Map of Kenya with the yellow dot showing Kiambu county where pigs and humans isolates were sampled

2.3 Antibiotic Susceptibility Testing

I determined antibiotic susceptibility test at KEMRI (September 2017) for 96 strains against a panel of 17 antibiotic drugs in three Mueller Hinton Agar (Oxoid, UK) plates per isolate using Kirby – Bauer Disk Diffusion. These drugs were; ampicillin (AMP), amoxicillin-clavulate (AMC), ceftazidime (CAZ), nalidixic acid (NA), gentamicin (CN), chloramphenicol (C), ciprofloxacin (CIP), cefoxitin (FOX), ofloxacin (OFX), trimethoprim-sulfamethoxazole (SXT), nitrofurantoin (F), linezolid (LZD), quinupristin-dalfopristin (QDA), amoxicillin (AML), imipenem (IPM), erythromycin (E), doxycycline (DO). Their diameter zone inhibition breakpoints results were interpreted with clinical and laboratory reference guidelines 2015.

2.4 DNA Extraction

The colonies that were identified to be *S. aureus* in step 2.2 were revived by streaking on Mueller Hinton Agar plates and incubating for 18-24 hours at 37^0C. Genomic DNA was extracted using the GenElute DNA kit (Sigma-Aldrich, USA) following the manufacturer's instructions. DNA were prepared ready for shipment on 96 barcoded wells according to the specifications of the WTSI sequencing facility. The DNA extraction and preparation for shipment of DNA were carried out by collaborators at KEMRI (December 2017).

2.5 Library Preparation and Whole Genome Sequencing

The WTSI core sequencing department prepared index-tagged multiplex paired-end libraries, and then sequencing of the samples were completed on the Illumina-B HiSeq X platform generating 150 bp paired-end reads, following previously described protocols (Quail et al., 2008, Harris et al., 2010b).

2.6 *S. aureus* Assembly and Annotation

Following Illumina sequencing of the isolates, the raw reads were *de novo* assembled and annotated automatically with the Sanger Pathogen Informatics pipeline as described by Page et al. 2016. In brief, multiple assemblies of the sequence data of each isolate were first generated using Velvet (v1.2) (Zerbino and Birney, 2008) and VelvetOptimiser (v2.2.5)(http://bioinformatics.net.au/software.velvetoptimiser.shtml). The SMALT(http://www.sanger.ac.uk/science/tools/smalt-0) was used to identify assembly reads that were mapped to different contigs or aligned to the same contigs but in improper orientation or that had different insertion sizes or were totally unmapped, to be subjected to further assembly improvement steps. This involved scaffolding of contigs of best N50 using SSPACE v2.0 (Boetzer et al., 2011) followed by closure of the gaps (1 or N's) with GapFiller v1.11 (Boetzer and Pirovano, 2012) that were generated during scaffolding of the contigs. The closure was done through cycling of BWA (Li and Durbin, 2009) and Bowtie (Langmead et al., 2009). The contigs that were shorter than the targeted fragment size length (300-500 bases) were removed in the final assembly. Finally, statistics of the assembly

quality were produced by mapping back the sequence reads to the final assembly using SMALT.

The final assembly sequences in FASTA format proceeded automatically to the annotation step that was performed using PROKKA (v1.11) (Seemann, 2014). PROKKA used PRODIGAL (Hyatt et al., 2010) to identify the coding sequences and compared them to genus specific databases from NCBI Reference Sequences (RefSeq) (Pruitt et al., 2012). The BLAST+ method (Camacho et al., 2009) was then used to search for the highest similarity results that were transferred for annotation

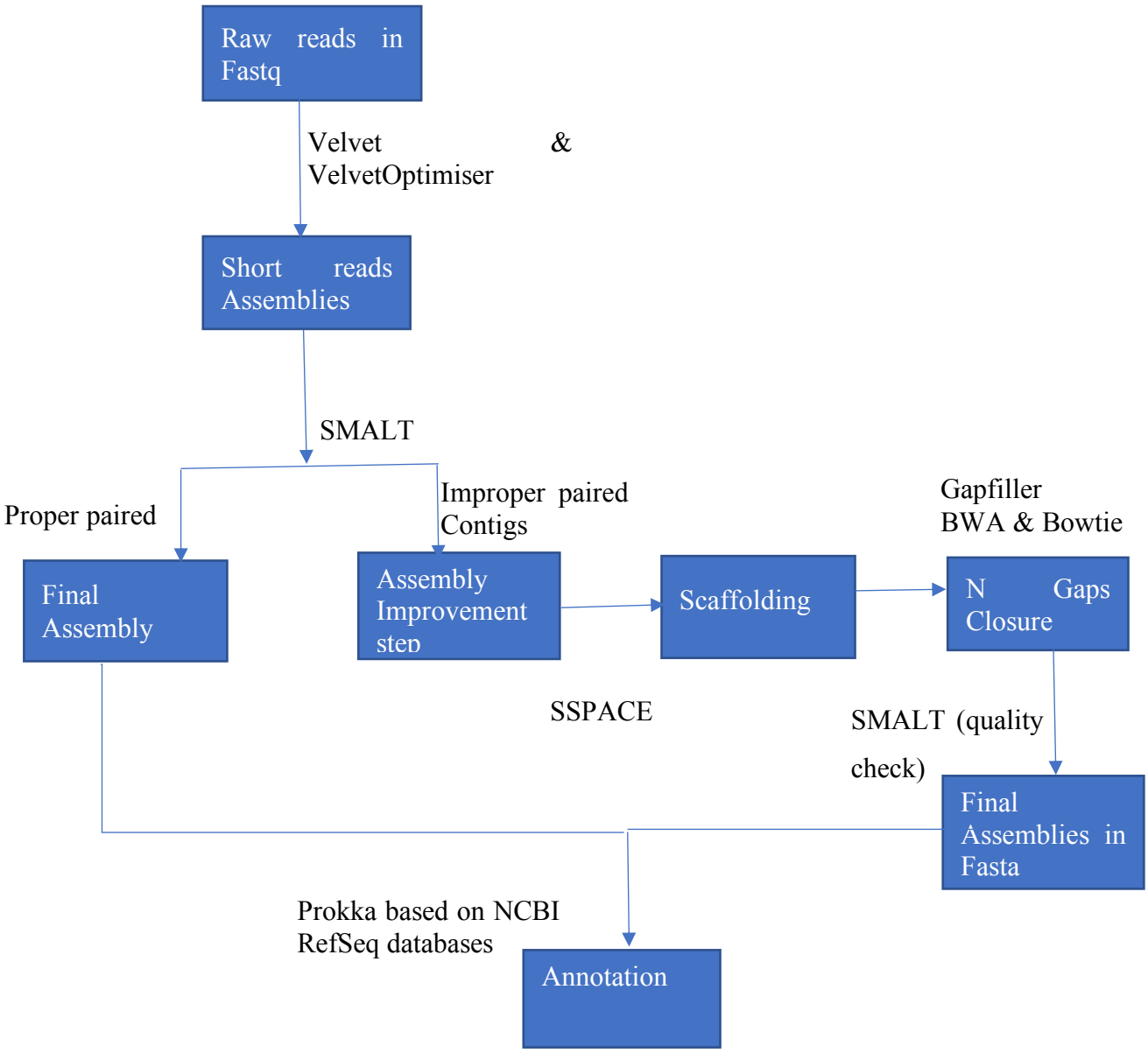


Fig. 2: A schematic diagram summarizing the assembly of *S. aureus* genomes by the Pathogen Informatics Pipeline as described by Page et al. 2016

2.7 Mapping and Variant Calling

2.7.1 Using Pathogens Informatics pipeline

Using the Pathogen Informatics pipeline, Illumina raw reads were mapped onto different reference genomes (table 1) depending on the analyses with Burrows Wheeler Alignment (BWA v0.7.15-r1140) (Li and Durbin, 2009). Briefly, alignment of the raw reads to the reference was done under default settings of BWA. After mapping, base call substitutions were identified using SAMtools mpileup v0.1.19 (Li et al., 2009) with Picard v1.92 detecting and marking duplicates in the BAM file. Bcftools v0.1.19 was used to filter out low quality SNPs that failed to meet the stringent threshold of the parameters in Table 2

Table 1: Different references used in mapping and SNP-calling analyses.

Dataset	**Reference strain (complete genomes)**	**Accession**	**Mapping pipeline**	**Identification of Recombination regions**
Kenyan isolates	*Staphylococcus aureus* subspecies aureus MSSA476	GCF_000011525_1	Pathogens Informatics	Identified by Mathew Holden
Global context of Kenyan isolates	*Staphylococcus aureus* subspecies aureus TW20	FN433596	Pathogens Informatics	Provided by Silvia Argimon
ST152 datasets	*Staphylococcus aureus* BB155	GCF_900004855_1	Custom script	Identified by Matthew Holden
ST188 strains	*Staphylococcus aureus* FORC_039	GCF_002140115_1	Custom script	Gubbins detection
Genetic comparison ST580 and ST398	*Staphylococcus aureus* subspecies *aureus*	GCF_001887075_1	Custom script	Gubbins detection for ST398 strains

	LA_MRSA_ST398			

Table 2: Quality control thresholds that variant sites must pass for them to be called SNPs

Parameter	**Thresholds**
Bcf variant quality score	>50
Mean Phred quality score	>30
Same allele frequency (af) of the base call as the reference	0
Allele frequency of SNP Variant base call of the reference	>0.95≤1
Read mapping of the base call	Ratio > 0.75
Mapping depth	> 4 reads
Mapping depth per strand	> 2 reads
Strand_bias, mapped_bias, tail bias	P value < 0.001

The pipeline produced VCF files that had unfiltered variants, filtered variants sites and pseudogenome alignment in FASTA format. A pseudogenome that was reconstructed had Ns in addition to SNPs, to represent the variant sites that failed filtering thresholds and bases that were deleted with respect to the reference. However, it lacked any bases that were inserted with respect to the reference.

2.7.2 Using a Custom Script

A custom script (developed by Simon Harris) employed a similar approach to the Pathogen Informatics pipeline in mapping and calling SNPs as described in step 2.7.1 such as use of BWA for alignment, SAMtools v-1.2 mpileup for detection of variants, BCFtools for quality control filtering of SNPs and Picard for marking duplications. However, the default settings for calling SNPs in this custom script are shown in Table 3

Table 3: Filtering parameters with cut-off values in the custom mapping and SNP-calling script.

Filtering parameter	**Cut-off value**
Minimum base call quality	50
Minimum mapping quality	20
Minimum depth reads matching SNPs	8
Minimum depth reads matching SNPs per strand	3
Minimum quality ratio of SNP/mapping reads	0.8

2.8 *In Silico* Prediction of Recombination Regions in Pseudogenome Alignments

Genealogies Unbiased By recomBination in Nucleotides Sequences (Gubbins) (https://github.com/sanger-pathogens/gubbins) (Croucher et al., 2015) was used with default parameters to predict recombination regions. These were identified as regions with a higher density of base substitutions in a pseudogenome alignment generated in step 2.7. In summary, phylogenetic reconstruction of the tree was performed with RAxML using the pseudogenome alignment, followed by reconstruction of SNPs on the phylogeny using Phylogenetic Analyses by Maximum Likelihood (PAML). This was followed by successive iterations of phylogenetic reconstruction with RAxML based on a reduced alignment after detection and subsequent removal of recombination events. At least 3 base call substitutions were used in identification of recombination regions. The output files include a newick tree file and the predicted recombination regions in GFF format, obtained upon detection of similar recombination regions on two successive repeats, and were visualized with the web-based application, Phandango (Hadfield et al., 2018).

2.9 Phylogenetic Reconstruction

Regions such as insertion sites, phages, transposons and mobile genetic elements defined in EMBL tab-delimited format were first removed from the pseudogenome alignment with a python script, *remove_blocks_from_aln* (https://github.com/sanger-pathogens/remove_blocks_from_aln). Variable positions that were exclusively 'ACTG' (without Ns) in all samples were extracted using the *SNP sites* tool (Page et al., 2016). Finally, the phylogenetic tree was inferred with RAxML-HPC v7.0.3 (Randomized

Axelerated Maximum Likelihood for High Performance Computing) using the SNP alignment. The RAxML-HPC used the general time reversible (GTR) model with a gamma correction for nucleotide substitution and 100 rapid bootstrap replicates. The resulting tree was mid-point rooted in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and exported to Microreact (https://microreact.org/) (Argimon et al., 2016), a web-based visualization tool, and combined with metadata.

In some instances, optimization of SNPs in the phylogeny and quantification of number of SNPs at the branches and the effect of SNPs on annotation were completed with a custom script (developed by Simon Harris). The script required pseudogenome alignment generated in 2.7 and best phylogenetic tree as the inputs for homoplastic parsimony, and the annotation of the reference in embl format for SNP effect prediction. It then performed optimization based on accelerated transformation (ACCTRAN) algorithms.

2.10 Determination of Antimicrobial Resistance and Virulence Genes using ARIBA

The presence of antimicrobial resistance (AMR) genes was determined directly from sequence reads using ARIBA (Hunt et al., 2017) with Resfinder (Zankari et al., 2012) and Arg-annot (Gupta et al., 2014) reference public databases, and custom databases of known AMR genes and mutations (developed by Matthew Holden and Sandra Reuter). In summary, coding sequences of the similar genes for the reference database were first clustered together using CD-HIT on default settings (Table 4) (Huang et al., 2010). This was followed by alignment of the paired raw reads of each strain to each cluster of the reference database with *Minimap* (Li, 2016). The reads that mapped to each cluster of reference sequences were assembled separately using Fermi-lite (https://github.com/lh3/fermi-lite). Then contig sequences of the assemblies with highest percentage to the cluster reads of the reference were identified with *nucmer* and variant information between the sequences and indels were detected with the *show-snps* program of the MUMmer package (Kurtz et al., 2004). In order to verify the completeness of the assemblies to the reference and identified known variants, the cluster reads were mapped back to the contigs with Bowtie2 (Langmead and Salzberg, 2012), and the depth coverage of the contigs and variants sites were determined with SAMtools mpileup (Li et al., 2009).

Heavy metals and antiseptic resistance protein genes sequences for the accession numbers in table 5, were downloaded from ENA and their presence among 23 strains were predicted in ARIBA as described before in AMR determination.

The virulence genes were predicted with a custom database of 102 virulence genes (provided by Matthew Holden and Ewan Harrison) using ARIBA as highlighted in the AMR determination.

Table 4: Default parameters in ARIBA.

CD-HIT clustering settings in preparation of reference databases	**Thresholds**
Sequence identity thresholds	≥0.9
Length sequence cut-off	0
Length of reference genes nucleotides	>6 ≤10,000
Number of genetic code to use	11
Number of threads	1
Nucmer default settings in selection of best match sequences	
Identity of sequence alignment	≥90
Length of sequence alignment	≥20
Depth coverage of reads for assembly	≥50
Number of reads for scaffolding of two contigs	≥10

Table 5: Showing heavy metals and antiseptic and their accession numbers

Heavy Metals and antiseptic	Accession numbers
arsB	AWW93856
arsC	PZH89330
cadX	AXE42895
cadD	PZL84176
qacA	AXE42913
qacB	BAG12275
qacC	SRE49840

2.11 Multi-Locus Sequence Typing (MLST) and *spa* Typing

The alleles of seven housekeeping genes of *S. aureus* used for multi-locus sequence typing (MLST), (https://pubmlst.org/saureus/), *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi* and *yqiL,* (supplementary file 1) were predicted *in silico* from assemblies using MLSTcheck script (https://github.com/sanger-pathogens/mlst_check). The nucleotide sequences of Protein A (*spa*) genes were extracted from the annotated assemblies using an in-house script (developed by Pathogen Informatics) and were uploaded to a web tool for assigning *spa* types (http://spatyper.fortinbras.us/). The tool uses Ridom and Kreiswirth nomenclatures for identification of repeat units and types of polymorphic X region of *spa* nucleotides.

2.12 Pan Genome Analyses of ST398 and ST580 strains

The 33 annotated assemblies of ST398 and ST580 strains produced as described in section 2.6 were used as input files (in GFF format) for Roary (Page et al., 2015), a pan-genome analysis tool. Roary was run using default settings, except that options -n to align each core gene family one by one with MAFFT and -e to generate a core genome alignment with PRANK. In summary, the initial step involved converting the nucleotide sequences in the annotated assemblies to protein sequences and subsequently filtering out the partial sequences. This was followed by iterative clustering of the remaining sequences with CD-HIT (Fu et al., 2012) starting with 100% identical sequences down to a default of 98% sequence identity. The last step involved pairwise comparison of each of the cluster sequences with BLASTP using the default parameter of 95% sequence identity and 100% matching sequence length. Subsequently, genes were then ordered and grouped into respective core or accessory regions based on their occurrence in the input sequences.

Genes that were present in more than 90% of the strains were discarded and the remaining 1863 genes were used for accessory genomes analyses. A python script was used to create a pairwise comparison of the number of accessory genes that were shared between any two strains of these 33 isolates. The pairwise comparison output was then converted to a matrix and subsequently clustered in R studio v1.1383 with ‘heatmap.2’ and coloured with ‘RColorBrewer’ from the ‘gplots’ package. The clustering was based on the proportion of shared genes in the accessory genomes between any two isolates.

2.13 Public Genomes Selection

117 genomes from a collection of over 10,000 *S. aureus* public genomes that were downloaded by Centre for Genomic Pathogen Surveillance from the European Nucleotide Archive (ENA) were randomly selected for inclusion in this study. In addition, seven genomes from the Washington National Primate Research Center in USA were downloaded from NCBI (Soge et al., 2016) using fast-dump package of the SRA toolkit program v2.92. The selection of genomes (Table 6) was based on the multi-locus sequence typing (MLST) results of the Kenyan *S. aureus* isolates from this study.

Table 6: Showed number of strains of public genomes from different study project numbers and the publications reference

No. of Genomes	Project Study No.	Reference
7	PRJEB12419	(Senghore et al., 2016)
18	PRJEB11627	(Strauß et al., 2016)
5	PRJEB12552	(Goncalves da Silva et al., 2017)
1	PRJEB12240	(Smith et al., 2016)
4	PRJEB12818	(Uhlemann et al., 2013)
1	PRJEB18560	(Edslev et al., 2018)
6	PRJEB2655 & PRJEB2944	(Jamrozy et al., 2017)
4	PRJEB2755	(Warne et al., 2016)
5	PRJEB2756	(Reuter et al., 2016)
3	PRJEB6236	(Kaas et al., 2014)
5	PRJEB7089	(Bletz et al., 2015)
2	PRJEB8084	(Mellmann et al., 2016)
15	PRJEB1915	(Moradigaravand et al., 2017)
11	PRJEB2096	(Holden et al., 2013a)
1	PRJEB2510	(Holden et al., 2013b)
18	PRJEB9575	(Moradigaravand et al., 2017)
6	PJNA306753	(Soge et al., 2016)
3	PRJEB9644	Permissions from Dr. Dorota Jamrozy
1	PRJEB11177	(Tosas Auguet et al., 2016)

6	PRJEB11281	(Nielsen et al., 2016)
3	PRJEB14187	(Julia Bünter 2016)

2.14 Best-Hit Genes Investigation

The genes of interest in Fasta format were individually compared to blast databases of nucleotides sequences of each assembly strain using an in-house python script (written by Sophia David) that employed NCBI Blastn (Altschul et al., 1990) for pairwise comparison algorithms. This produced the best hits (100% identical) genes presence and their locations in the genome for easy identification in ARTEMIS visualization. In some cases, the nucleotides sequences of the best hit genes were aligned with muscle and exported to SeaView v4.3 in order to identify mutations and truncations.

2.15 Quality Control Checks

Statistics generated by Pathogens Informatics Pipeline were used in filtering out contaminated strains. They were produced based on mapping of raw reads to TW20 reference, de novo assemblies and annotation as described in step in 2.6, and Kraken pipeline for assigning reads to taxon on mapping to bacterial databases of NCBI RefSeq.

2.16 Study Project Number

The genomes of this study together with their metadata will be deposited in European Nucleotides Archives (ENA) (www.ebi.ac.uk/ena/) under project number ERP105373.
The results for the analyses for the 23 Kenyan isolates together with phylogenetic tree are found in microreact (https://microreact.org/project/B1WId9zXQ). The results of 23 isolates in combination with public genomes are found in this microreact link (https://microreact.org/project/rk0AaZBjz).

CHAPTER THREE

RESULTS

OVERVIEW

This chapter is sub-divided into two sections. The first section describes the results of quality control analyses, phylogenetic investigation as well as molecular characterization of Kenya isolates. The second section highlight the phylogenetic relationship of Kenya isolates in combination with 126 public genomes and correlation of certain genes with phylogenetic lineages.

SECTION A

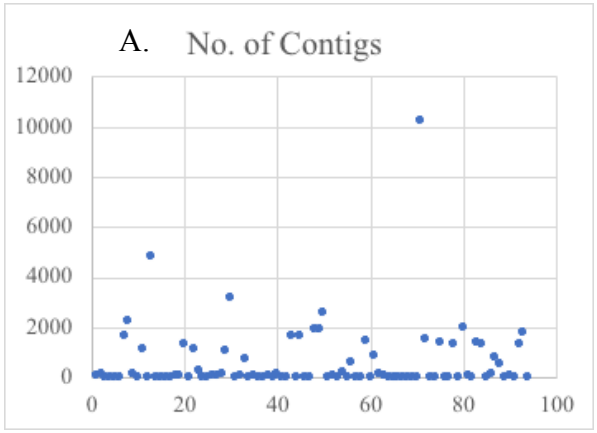
3.1.1 Study Population

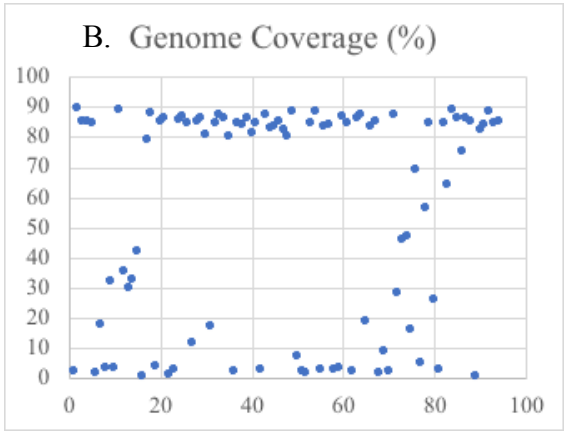
Ninety-four strains that were successfully sequenced, were randomly selected from the collections of samples from a study undertaken in Kiambu county, Kenya between October 2015 and August 2016 (unpublished data). These 94 sample isolates were obtained from nasal swabs of 46 humans and 48 swine in 29 different homesteads. Only up to a maximum of 3 farmers and 3 pigs from an average of 11 herd size (range 2-125 pigs per farm) were selected for sampling in the same homestead. All the pigs sampled were reared under zero grazing on a mean average of 2 acres piece of land with exceptions of 8 swine isolates that were grazing freely.

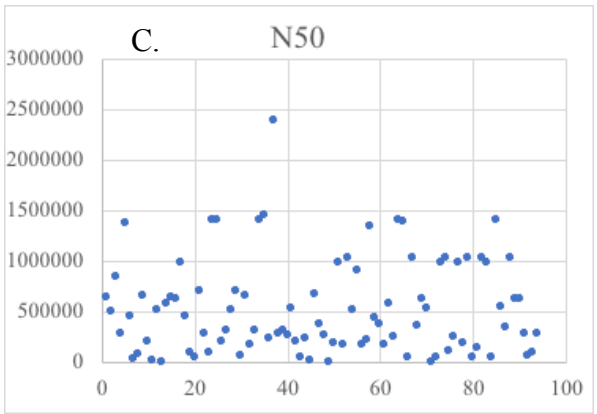
3.1.2 Genomic Metrics

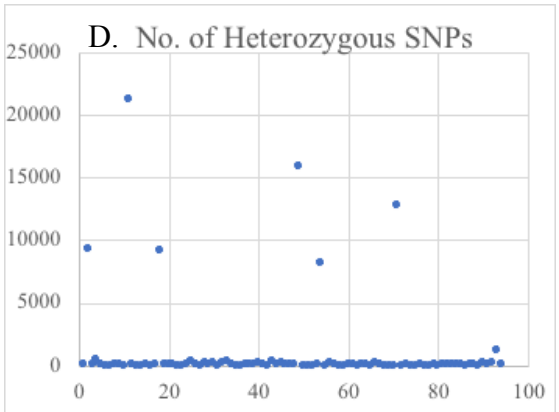
To ensure that genomes used for downstream analyses were of good quality, 94 sequence data were initially evaluated by plotting comparison graphs (as shown in fig 3) using different indicators of contamination such as number of contigs, N50, assembly length, genome coverage and number of heterozygous Single Nucleotides Polymorphisms (SNPs). Then rationale parameters (see table 7) were set for stepwise filtering of the suspected contaminant sequence strains. First, 60 isolates were eliminated because of low genome coverage on mapping to TW20 reference genome FN433596 and were either mixed with other bacterial species or different species as shown in Fig. 4. Next step was exclusions of

four isolates which had assembly length out of this range (2.6 to 3.6 Mbs). These genomic size thresholds were set based on estimation reported in the genome assembly and annotation of *S. aureus* in the NCBI (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/genome/genomes/154?). Among 30 remaining isolates, 4 were suspected to be contaminated because they had larger number of contigs (544-857) than set maximum (400). The 26 genomes were further screened based on the number of heterozygous SNPs (maximum n=1000) in which 3 genomes were excluded because this could suggest the genomes might have been mixed with other *S. aureus* strains. To rule out that these SNPs were arising possibly from prophages, we checked manually their distribution by visualizing their BAM files with ARTEMIS. As a consequence of applying the above thresholds, genomes with low depth coverage, large number of unmapped reads, high number of coding genes were also eliminated resulting in 23 genomes for downstream analyses and 71 isolates being excluded for further analyses.









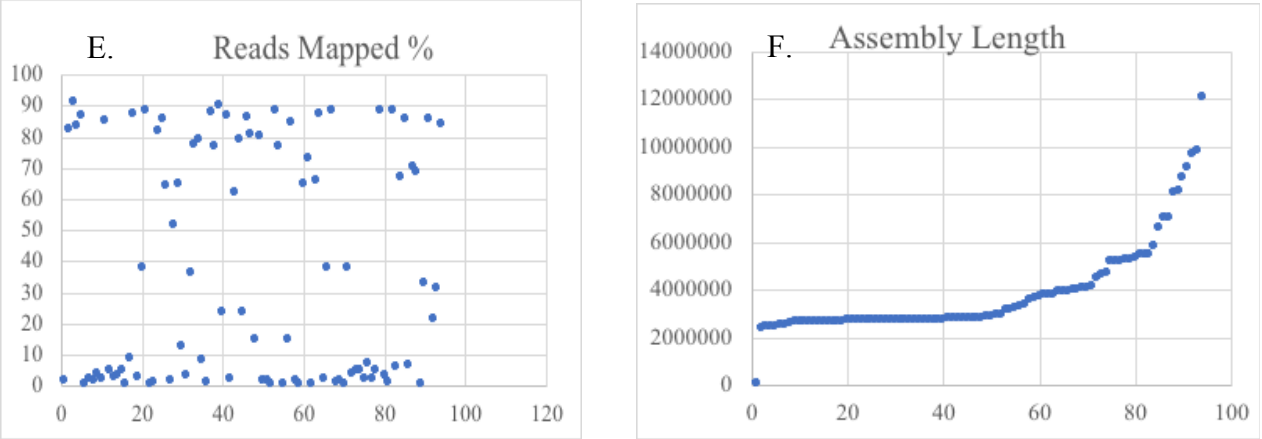


Fig. 3: Graphs showing distribution of 94 genomes based on number of contigs (A), genome coverage (B), N50 (C), number of heterozygous SNPs (D), Reads mapped % (E) and genome size length (F) before the QC filtering parameters were applied.

Table 7: Quality control filtering parameters

Filtering Parameter	**Threshold**	
	Minimum	**Maximum**
Genome size length	2.4 Mbs	3.6 Mbs
Number of contigs	N/A	400
Genome coverage compared to reference (%)	70	100
Number of Heterozygous SNPs	N/A	1000

The 23 *S. aureus* isolates have the following genomic metric features

Table 8: Genomic metrics of 23 high-quality genomes

Parameter	**Minimum**	**Maximum**	**Median**
Genome size length (bp)	2674035	2831812	2739599
Number of contigs	13	106	17
% Genome coverage to reference	85.32	96.05	93.94
Number of Heterozygous SNPs	39	401	115
Number of CDS genes	2420	2602	2518
Number of Genes	2549	2738	2648

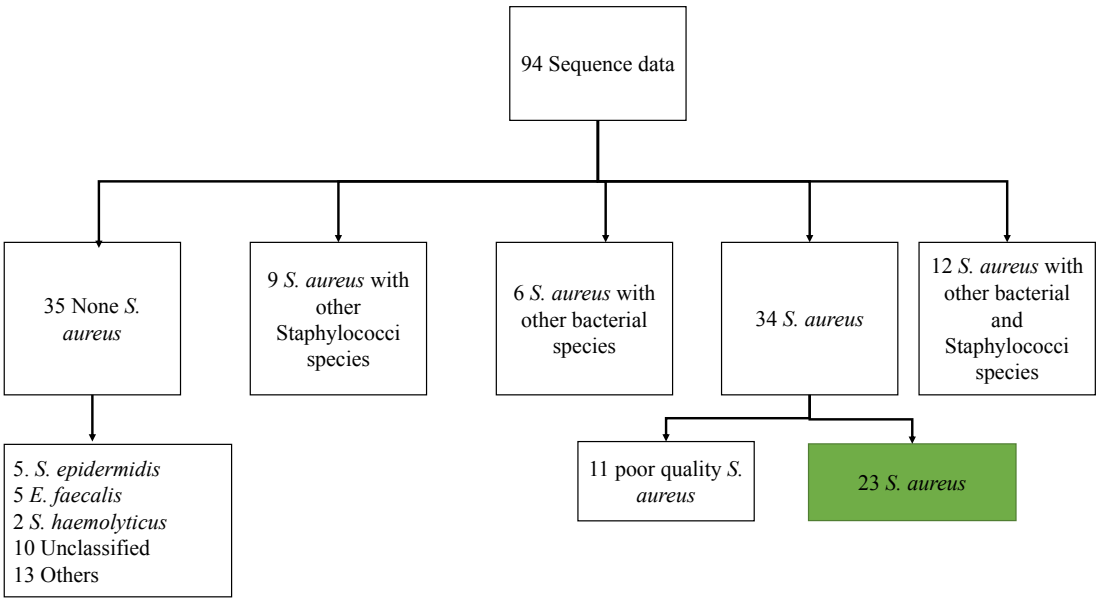


Fig. 4: A schematic diagram of a summary of general flow description of species identification and contamination of sequence data

Among 23 *S. aureus* high-quality genomes, 11 were sampled from swine and 12 from humans. Of these, 12 strains were paired in the homesteads (table 9). These paired isolates from the same homestead were sampled at the same time. The remaining 11 isolates (6 from swine and 5 from human) are un-mated and are distributed in 11 different farms. The average age of the pigs were 9 months while that of farmers were 25 years (range, 12 to 43 years). The range of herd size of pigs per farm were between 6 and 125 and were carried out under zero grazing (except in two homesteads that were grazing freely) in an approximately 2 acres piece of land. Only 3 farmers and none of the pigs had received antibiotics within the last 3 months prior to their sampling date (supplementary file 2, provide information of the farmers and pigs).

Table 9: showing pairing of strains in the homesteads (No. indicate number of strains while id. means serial identity of the homestead)

Host pairing	No. of strains	Homestead id.
Swine – human	6	24, 6, 33
Human – human	4	5, 13
Swine – swine	2	21

3.1.3 *In Silico* Prediction of Multi-locus Sequence Types and *spa* typing

To determine Multi-locus Sequence Types (MLST) of the strains, MLSTcheck script was used to compare sequence assemblies of these genomes against alleles of seven housekeeping genes in PubMLST reference database. The analyses of sequence types (ST) revealed a significant degree of genetic diversity in terms of clonal lineages in Kiambu county, Kenya. Among the 9 STs (Fig. 5A) identified in the 23 isolates, the dominant clone was ST188 26% (n=6) the predominant clone of livestock and hospitals in Pan-Asia continent (Wang et al., 2018). This was followed in frequency by ST789, ST580 and ST25 with 3 isolates (13%) in each ST. The ST188 is a double locus variant (DLV) of ST1, ST789 is a single locus variant of ST7 while ST580 is a DLV of livestock-associated ST398, a dominant clone in Europe and North America. These dominant population clones were shared among human and swine isolates (Fig. 5B).

In order to find out the distribution of Protein A (*spa*) types among the hosts, an in-house script was used to extract *spa* nucleotides from the annotated assemblies for determination of repeat units and types using Ridom and Kreiswith nomenclatures. However, the prediction of the *spa* type based on assemblies may not be accurate since the assigning of *spa* type number depend on tandem sequence repeats of polymorphic X region which could have been affected by assembly and sequencing errors. Nine *spa* types were predicted among 23 strains and were distributed among swine and human (fig 6) as defined in MLST results. The four 60% most common *spa* type of isolates were t189 (n=6), t7817 (n=3), t1176 (n=3) and t091(n=3). Overall, *spa* type and MLST results could be indicating higher diversity of clonal lineages circulating in Kiambu and their presence in both human and swine suggested possibility of these clonal lineages colonizing different types of hosts.

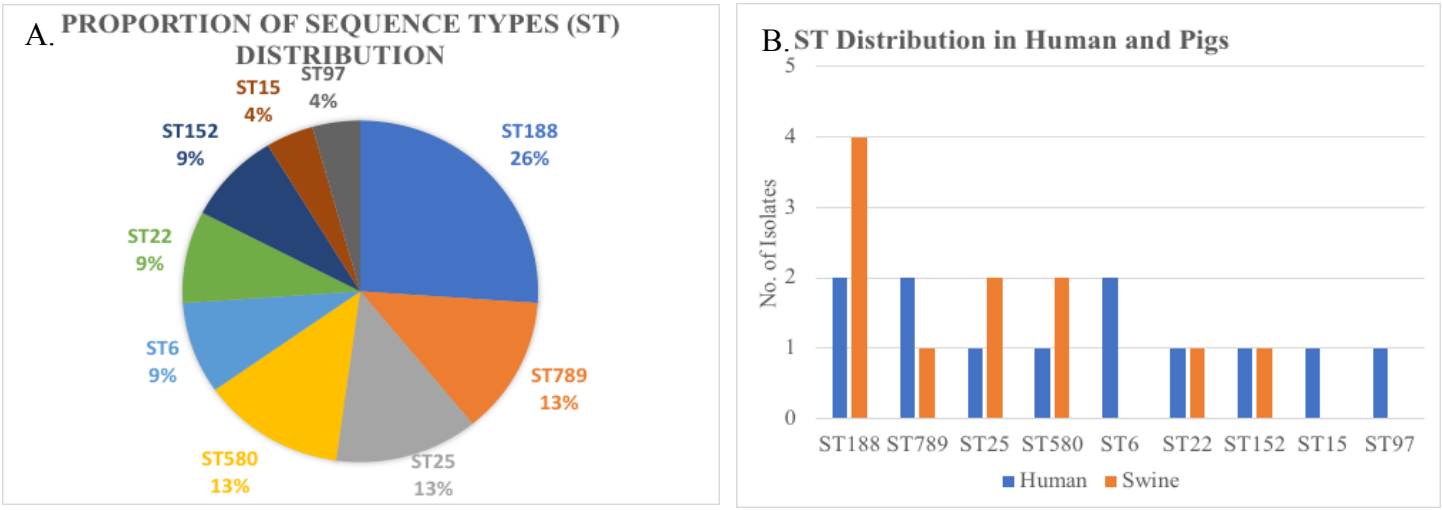


Fig. 5: Prevalence of STs (A) and distribution of STs per host (B)

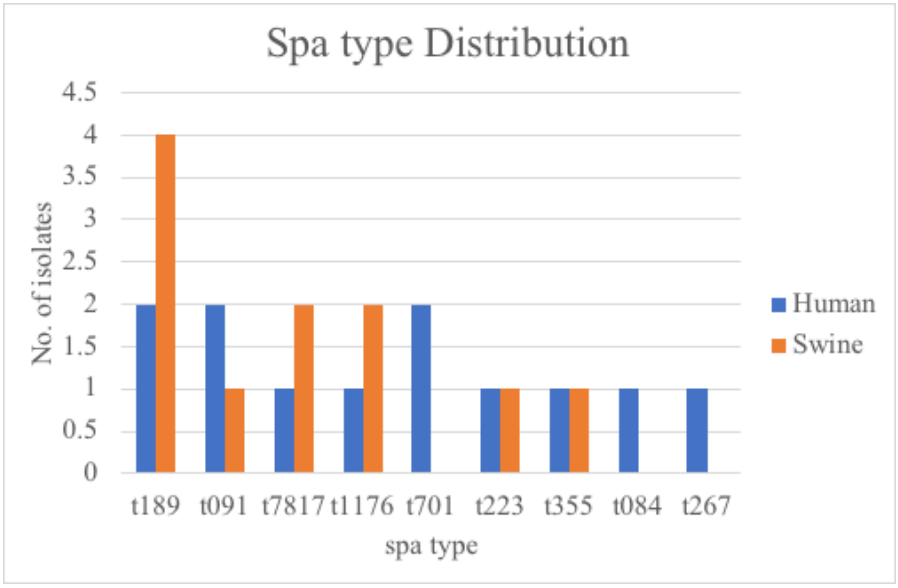


Fig. 6: *spa* type distribution between humans and swine.

3.1.4 Phylogenetic Analyses

In order to reconstruct phylogenetic tree for genetic diversity investigation between swine and humans in this study, an in-house pipeline (developed by Pathogens Informatics) was first used to retrieve a pseudogenome alignment of 2.82 Mb in length (without reference) that was generated by mapping raw reads to MSSA476 reference (ST1 by MLST) and subsequently calling SNPs. The 20 regions (185 kb size) of mobile genetic elements of the reference such as indels, prophages, transposons and other repetitive elements were removed from the alignment. This resulted in a concatenated core genome alignment of 2.63 Mb size that was used for SNPs extractions. The comparison of this concatenated core alignment of all 23 *S. aureus* strains identified 57010 variants that were exclusively ATCG, further supporting higher genetic diversity between clonal lineages of the 23 isolates. The phylogeny formed distinct clades according to ST and *spa* type designation with humans and swine isolates within and across homesteads intermingling together in the tree.

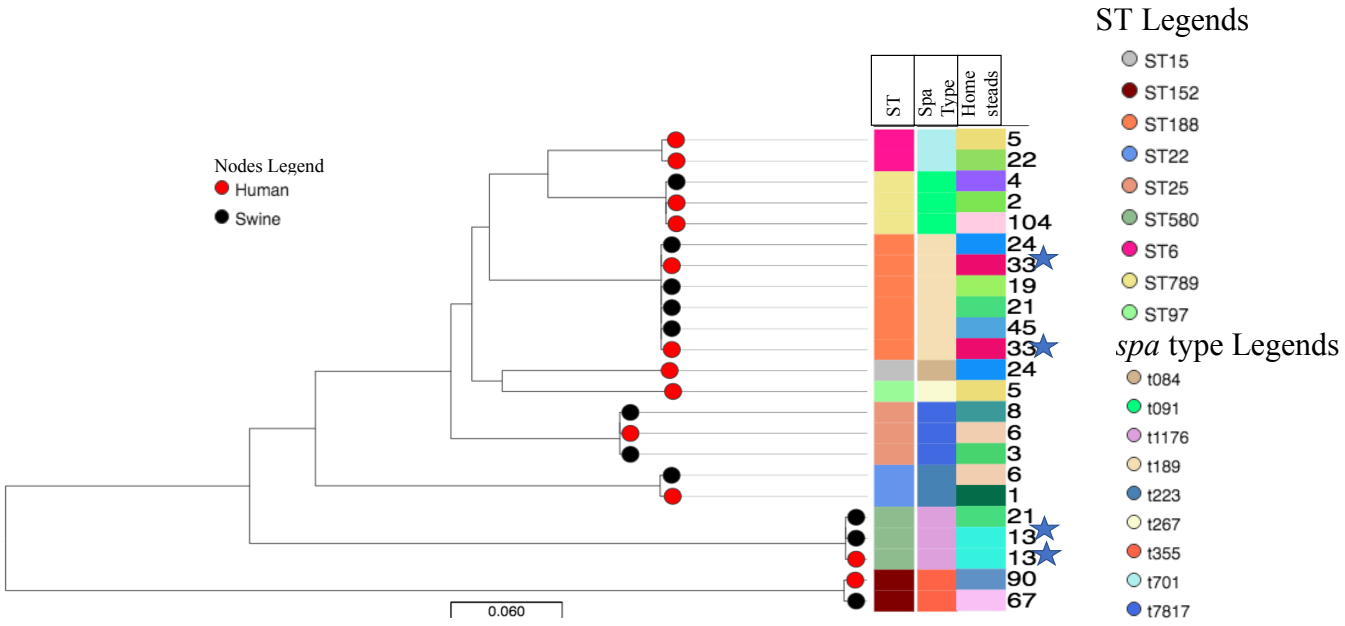


Fig. 7: Midpoint rooted maximum likelihood phylogenetic tree based on 57010 core genome SNPS. The source (human and swine) in the phylogeny nodes are linked to ST and spa type and homesteads number

Four pairs of strains (fig 7) sampled from the same homestead on the same date (i.e., homestead number 5 with human-human, 6 with swine-human, 21 with swine-swine and 24 with swine-human), belonged to different STs and had large number of SNPs differences between 8796 and 21793. On the other hand, two pairs of strains swine/human and human/human collected in homesteads 33 and 13 (blue star in fig 7), respectively, belonged to the same ST and had near identical genotypes with 2 and 6 SNPs differences, respectively. The genomes of humans and pigs within clonal lineages ST789, ST25 and ST188 were highly similar, with less than 5 SNPs differences, even though isolates were sampled from different homesteads (exact GPS locations of homesteads and date of sampling not provided).

The typing methods such as *spa* typing, are routinely used to infer transmission of *S. aureus* strains between humans and animals for epidemiological surveillance studies (Harris et al., 2010b). However, strains within each cluster of *spa* type t701, t223 and t355 had SNPs difference of 222, 86 and 68 respectively. Using less than 50 SNPs difference as a cut-off threshold for defining suspected recent sharing of common ancestor as determined by Coll et al. in longitudinal surveillance studies of MRSA in UK (Coll et al., 2017) indicate these isolates were distantly related.

3.1.5 *In Vitro* Antibiotic Susceptibility Test Results

The *in vitro* antibiotic susceptibility distribution profiles of the 23 strains were examined using Kirby – Bauer Disk Diffusion and their zone diameter breakpoints were interpreted with Clinical & Laboratory Institute Reference guidelines (supplementary file 3). Both isolates of humans and swine showed high prevalence of ampicillin (n=21, 91%), followed by erythromycin 40% n=9. Linezolid and chloramphenicol resistance were only detected in humans while one strain of swine exhibited non-susceptibility to ciprofloxacin. Only 3 strains were phenotypically resistance to cefoxitin (MRSA) (fig 8).

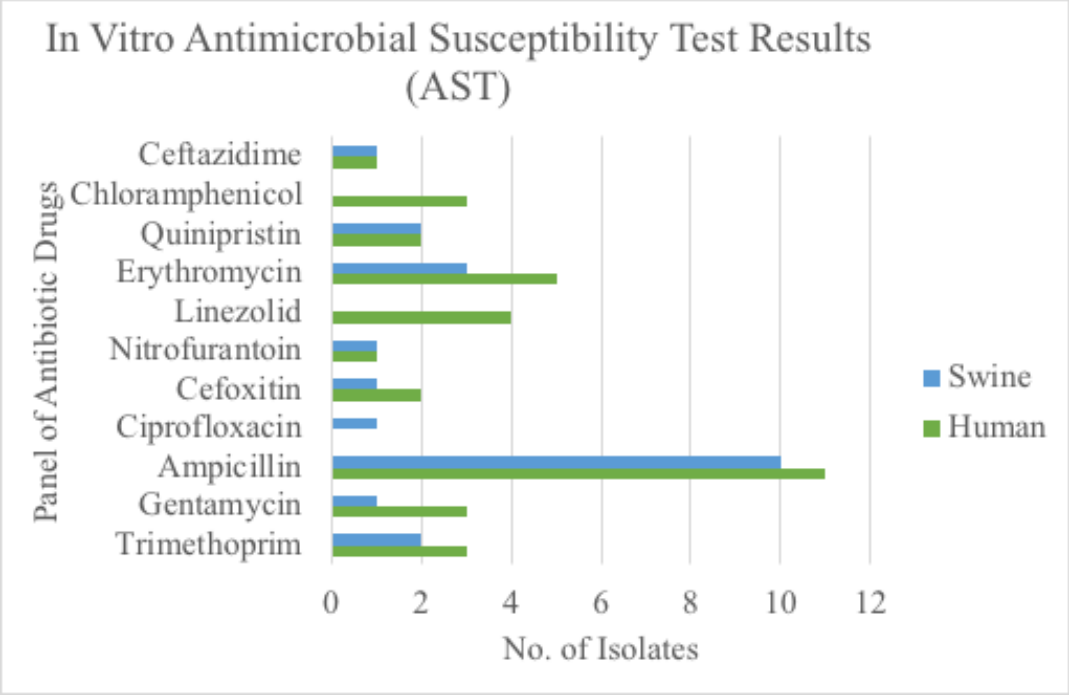


Fig. 8: Histogram showing phenotypic distribution of antibiotic drugs between pigs and humans

3.1.6 *In Silico* Prediction of Antimicrobial Resistance Genes

To investigate antimicrobial resistance (AMR) genes among the 23 WGS raw reads, reference databases of public resource Resfinder (Zankari et al., 2012) and Argannot (Gupta et al., 2014) and custom databases of known variants regions associated with resistance to fusidic acid, fluoroquinolones, vancomycin, rifampin, daptomycin and muciprocin were determined with ARIBA.

The resistance genes that were identified in 23 strains encode for aminoglycosides (*aadC*, *aac6_le_APH*), tetracyclines (*tetM*, *tetK* and *tetL*), β-lactams (*mecA* and *blaZ*), fosfomycin (*fosA* and *fosB*), trimethoprim-sulfamethoxazole (*dfrG*, *dfrC*, and *dfrK* variant) and macrolide-lincosamides-streptogramins (*mphD*, *ermC*, *lnuA*, *str*) (as shown in fig 9).

Generally, there was low level prevalence of AMR. For instance, known variants associated with resistant to ciprofloxacin, muciprocin, fusidic acid, daptomycin, vancomycin and rifampin were not detected in these isolates. However, novel variants and frameshift changes as a result of deletion were detected in *gyrA* at T818 in ST580 and ST789 strains (Supplementary File 4).

The resistance gene *blaZ* which encodes for penicillin resistance was the most prevalent (22/23, 95%). Genes that were significantly identified in higher frequency are those which encode fosfomycin, trimethoprim and macrolide each with 4 isolates (17%) represented by *fosB*, *dfrG* and *ermC* genes respectively. Only one strain carried *mecA* gene that confer resistance to methicillin. Two ST789 isolates (one swine and human) were found to be multidrug resistant and carried *blaZ*, *tetL*, *str* and *lsaa* genes, and one human ST25 isolate that had blaZ, dfrG, fosB, ermC and tetK genes. There was no significant difference in the distribution of classes of antimicrobial resistance genes among the swine and human isolates (Fishers Chi^2 exact test ($p=0.1845$) p-value < 0.010).

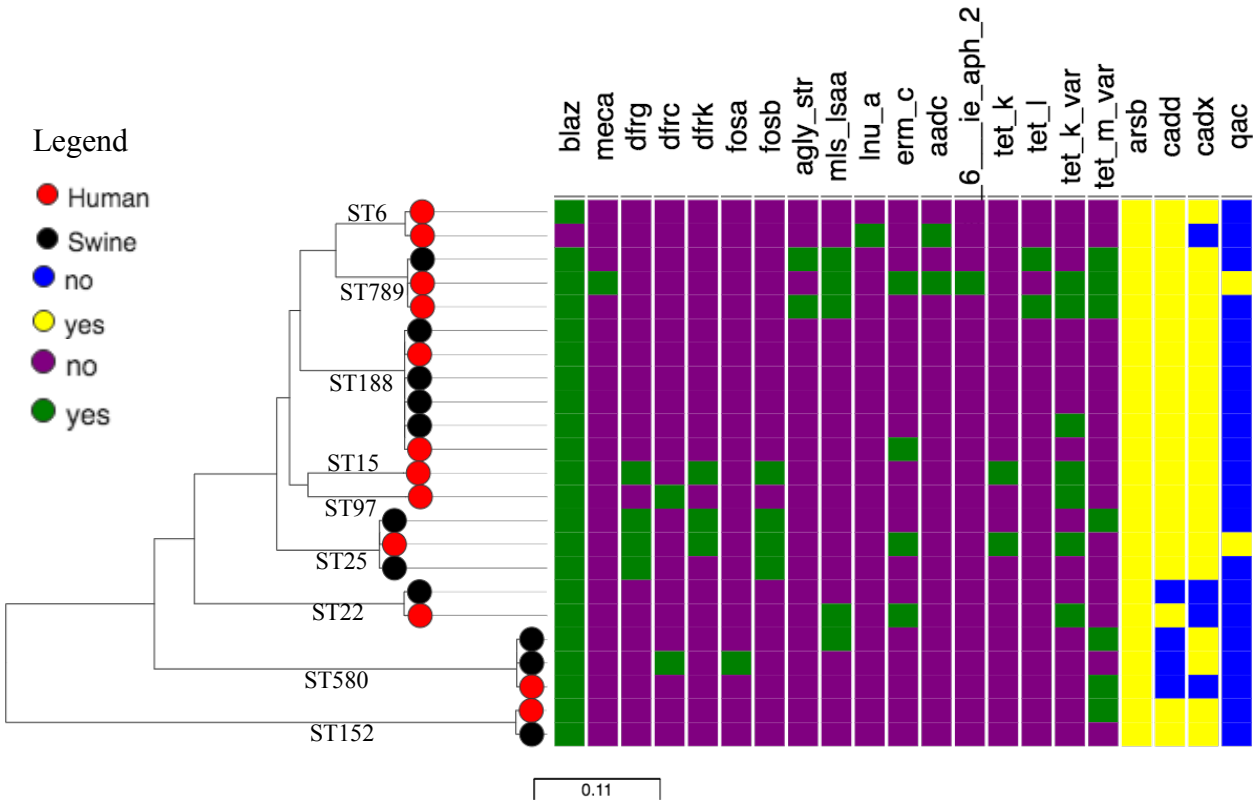


Fig. 9: A midpoint rooted maximum likelihood phylogenetic tree with heatmap distribution of AMR genes and heavy metals. The names at the top of heatmap indicate AMR and heavy metal gene names

The comparison of *in silico* prediction and phenotypic resistance available in supplementary file 5 based on the commonly used antimicrobial agents in the treatment of *S. aureus* infections as suggested by (Aanensen et al., 2016b, Gordon et al., 2014) in their studies for determination for the usefulness of whole genome sequencing in prediction of AMR. The results may not be accurate since the single colony used for antibiotic susceptibility test could not have been used for DNA extraction for subsequent sequencing.

3.1.7 *In Silico* Prediction of Virulence Genes

The presence or absence of virulence genes among 23 WGS data were predicted with ARIBA using a custom database of 102 set of virulence genes as the reference with a minimum threshold of 95% nucleotide identity and 99% coverage of query contig length to predict presence of a gene. A variety of virulence genes were identified in at least one of the strains encode for enterotoxins, leukocidin, capsular 8 polysaccharide, immune evasion genes clusters, haemolysins, biofilm formations, iron proteins, toxic shock syndrome, arginine catabolic metabolite enzymes (ACME),adhesive and serine proteases genes (fig 10).

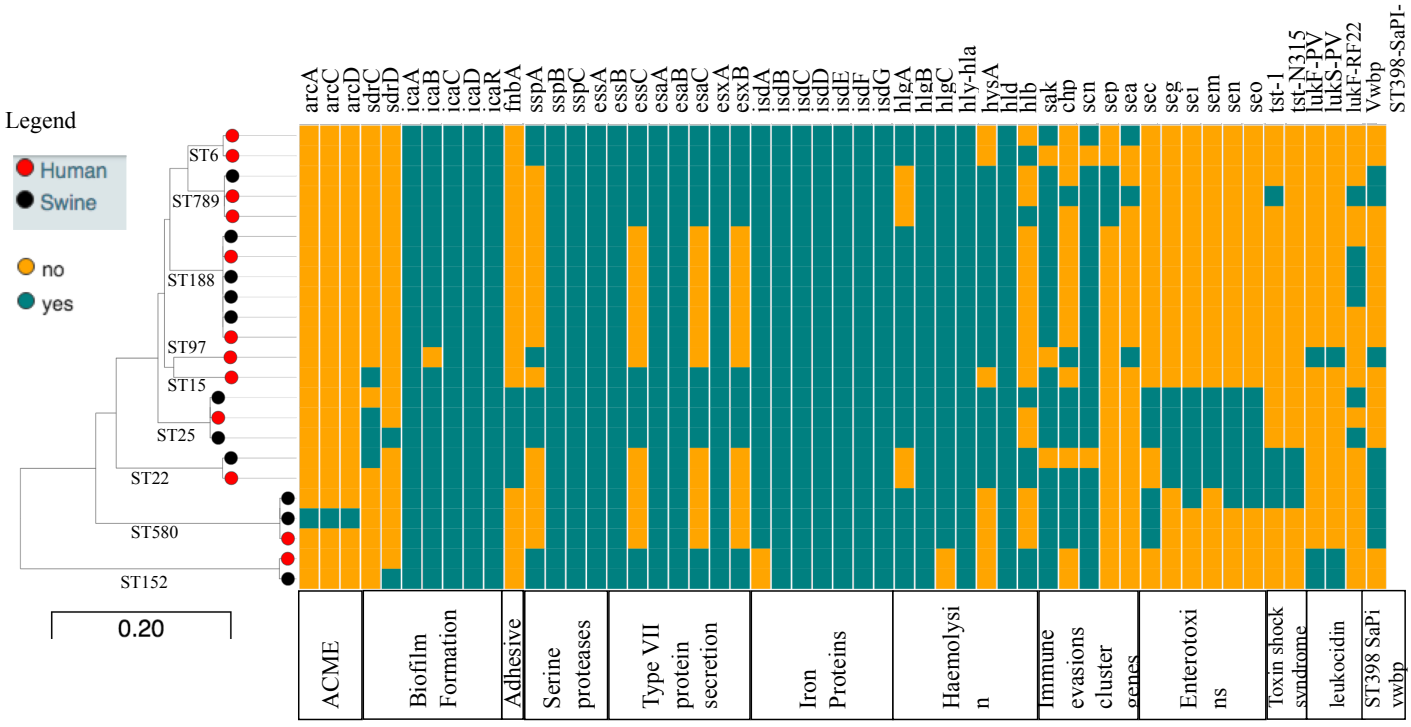


Fig. 10: A Midpoint rooted phylogenetic tree with a heatmap of virulence genes (not included capsular 8 polysaccharide) that were present in at least of one strain of the 23 isolates. Nodes indicate source of the isolates and genes names indicated at the top and the broad classification of the genes at the bottom

The Panton- Valentine leucocidin (PVL), encoded by two sub-unit proteins, lukF-PV and lukS-PV which is associated with skin and soft tissues infections such as necrotizing dermatitis (Lina et al., 1999) was present in 13% (n=3) strains including one isolate from swine origin of ST152.

Staphylococcal food poisoning characterized with emetic activity and nausea is likely to occur due to consumption of food contaminated with *S. aureus* which had been improperly cooked or stored allowing for their growth and expression of enterotoxins (Hennekinne et al., 2012). The classical enterotoxins (SEA-SEE) have been commonly implicated in food poisoning. About 15% (n=3) of human and none of the swine isolates carried the *sea* genes, and nearly a quarter (4 from swine and 2 from human) of the strains carried *sec* genes.

The enterotoxin gene cluster (*egc*) that comprises of *seg, sei, sem, sen* and *seo* genes were the most prevalent 6 (26% of the isolates) among the novel enterotoxins and were found mostly in strains of ST25 and ST22 lineages. To determine their genetic arrangement in each lineage, an in-house Blastn script was used to identify presence and location of these genes in the ordered assemblies and visualized their annotated GFF format file in Artemis (as shown in the figure 12). These lineages have similar genetic arrangement of these *egc* except presence of enterotoxin sec-like 1 (*entC1*) in ST25 and sec-like 3 (*entC3*) in ST22.

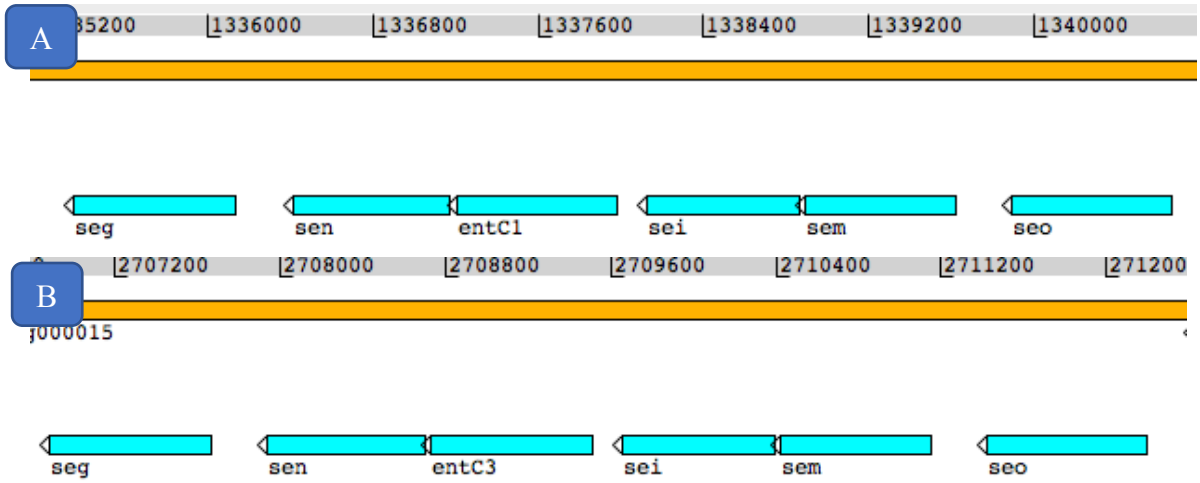


Fig. 11: Artemis visualization of enterotoxins gene clusters in ST25 (A) and ST22 (B)

Genes *tst-1* that mediate for toxic shock syndrome in humans were identified in 4 (17%, 2 from humans and 2 from swine) isolates, and 3 were co-detected with *egc* genes.

The distribution of human immune evasion gene clusters (IECs) among human and swine isolates was further explored. These genes are found in prophages of integrase group 3 (*phi*3) that are integrated into the ß-hemolysin genes (*hlb*), and are highly specific for human neutrophils and therefore, their presence in the host's isolate could indicate human origin.

Approximately 90% (n=21) of the 23 isolates tested for immune evasion gene clusters (IECs) were positive for two or three of the IECs genes (table 10). This suggests that about 90% (n=10) of the swine isolates in this study were colonized by isolates that were likely to have originated from humans. Only two isolates (ST22 from swine and ST6 from human) belong to phage type H that are associated with livestock adapted lineages.

Table 10: Different types of IECs (van Wamel et al., 2006b) with their prevalence among 23 strains

IECs type	Genes	Overall Prevalence	Human (n=)	Pigs (n=)
Phage type E	*sak* and *scn*	39%	4	5
Phage type B	*sak*, *chp* and *scn*	30%	3	4
Phage type D	*sea*, *sak* and *scn*	4%	1	0
Phage type H	lack all the IECs genes	8%	1	1
Not-yet typed	*sea*, *sep*, *sak*, *chp* and *scn*	4%	1	0
Phage type G	*sak*, *scn* and *sep*	8%	1	1
Not-yet typed	*sea*, *chp* and *scn*	4%	1	0

SECTION B

3.2.1 Relatedness of Kenya Isolates with Publicly Available Genomes

To contextualize humans and swine *S. aureus* genomes of this study in global genetic perspective, their genetic relatedness was compared with 126 publicly available genomes drawn from diverse countries as guided by MLST results but with a major focus on African genomes (supplementary file 6). The ST distribution of these public genomes is as follows: ST152 (n=25), ST188 (n=24), ST6 (n=14), ST22 (n=12), ST25 (n=19), ST398 (n=30) and ST789 (n=2). The public genomes of two STs (ST97 and ST15) that had only single strain in each were not included and the search for literature for public genomes in PubMed for ST580 were missing.

First, the genomes of 105 randomly selected genomes and 23 Kenyan isolates were mapped to TW20 complete genome to identify core genome SNPs (completed by Pathogen Informatics pipeline). After excluding 367 kb size of associated regions of MGEs of the reference in the alignment, 42851 core genome SNPs were identified among 128 isolates. This core genomes SNPs was used for maximum likelihood phylogenetic reconstruction with RAxML.

Generally, the SNPs diversity of isolates of each ST were increasing over time in terms of SNPs pairwise differences on mapping to TW20 reference (Fig. 12).

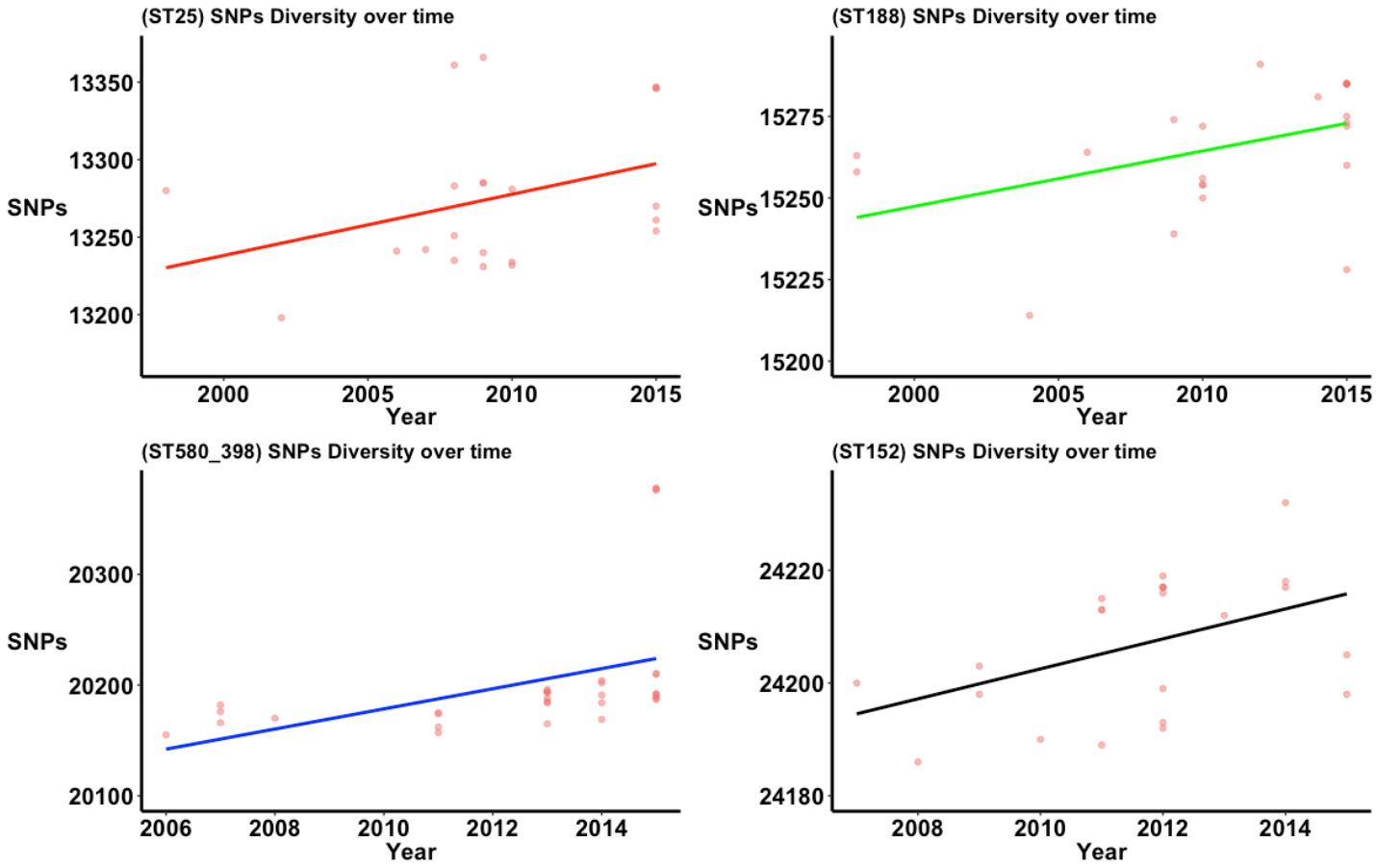


Fig. 12: Regression lines of each ST in terms of SNPs pairwise difference over time (in years) in relation to TW20 (ST239). The dot represents the number of SNPs of individual strain while the line represents the line of best fit distribution of isolates generated in ggplot2 in RStudio.

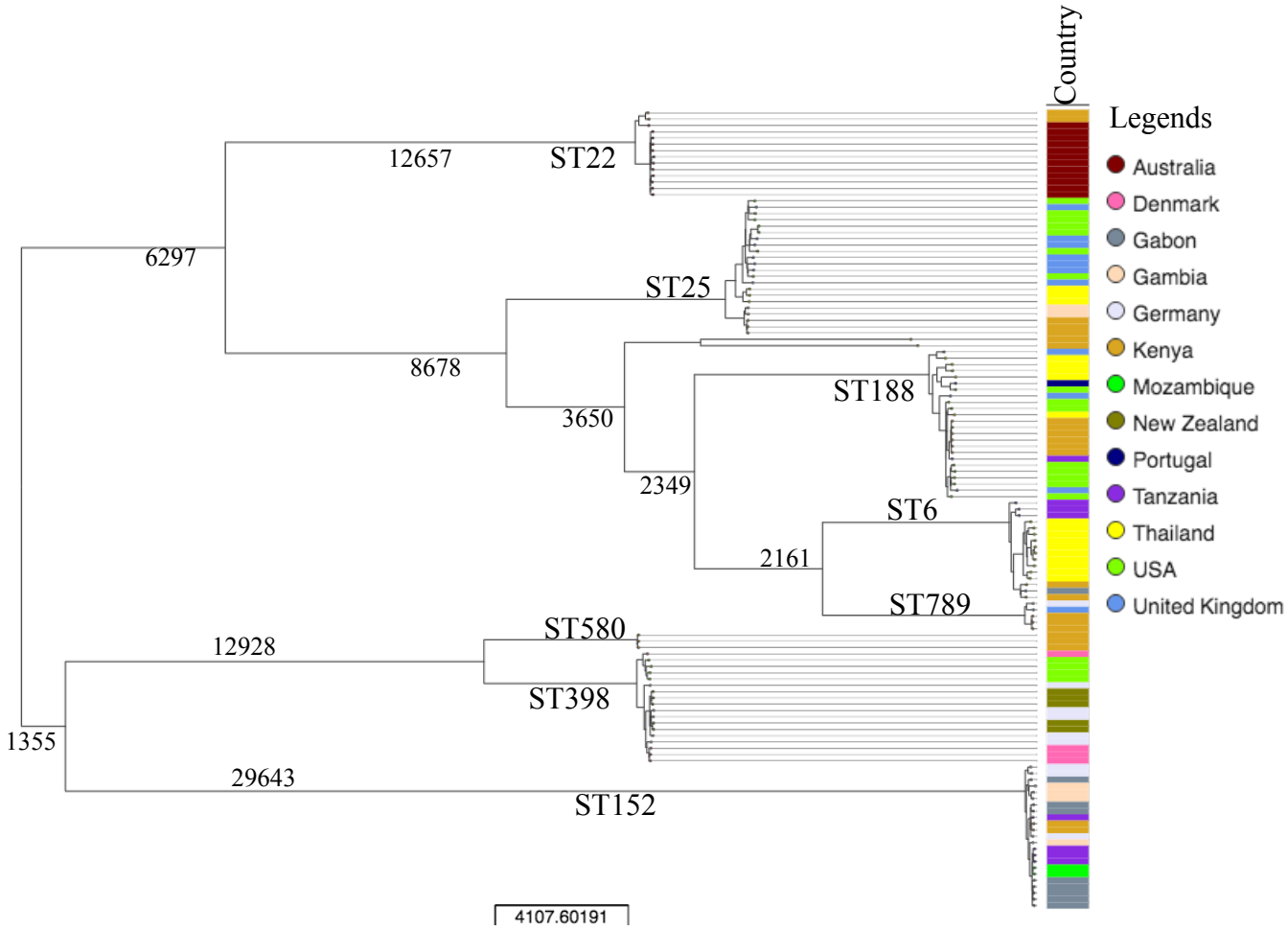


Fig. 13: A Midpoint rooted maximum likelihood phylogenetic tree that has been optimized with ACCTRAN algorithms. The tree tips are aligned to country of origin. The legend indicates the color of country. The number in the branches indicate number of SNPs in the branch

A midpoint rooted maximum likelihood phylogeny (fig 13) segregated into MLST lineages to form the same 7 distinct clusters as shown before in phylogenetic tree of Kenya isolates (fig 7) suggesting associations of lineages circulating in humans and swine isolates of Kenya with these global lineages. Strikingly, ST152 clusters were characterized with long deep branch as compared to other MLST clusters suggesting that they are distantly related to other clusters.

3.2.2 Detailed Phylogenetic Analyses of Selected Sequence Types (STs)

To better understand the genetic clustering of Kenyan isolates in relation to global genomes, individual subtrees of individual ST (ST25, ST789, ST22, ST6) were reconstructed based on their respective core genome SNPs on alignment to TW20 reference. Nevertheless, ST188 and ST152 were mapped to respective ST references using custom script. Fourteen regions of 132 kb size that were associated with repetitive elements of the reference were excluded from the alignment in ST152 resulting in 1446 core genome SNPS for maximum likelihood reconstruction of phylogeny. To examine the effect of these core genome SNPs in the genomes of ST152, accelerated transformation was used to reconstruct maximum likelihood phylogeny with pseudogenome alignment and annotated reference file. This led to identification of non-synonymous, synonymous SNPs and intergenic (table 11).

Table 11: shows SNPs effect of core genome of ST152 as determined by ACCTRAN reconstruction of phylogeny with original pseudogenome alignment and annotated reference file

No. of SNPs	Type of SNPs
512	Intergenic
990	Synonymous
1066	Nonsynonymous
23	Non-stop codon to stop codon

The recombination regions were predicted using Gubbins in ST188 where it identified approximately 71kb size repetitive regions of the pseudogenome alignment (fig 14) and were removed before SNP sites extraction in genome alignment and RAxML phylogenetic tree reconstruction. These recombination regions were phages with various genes that annotate for functions such as phage proteins of unknown functions, phage major capsid, putative phage PVL protein, phage tail fiber proteins, hypothetical phage related proteins, putative phi ETA-like proteins, bacteriophage protein of unknown functions among other phage functions.

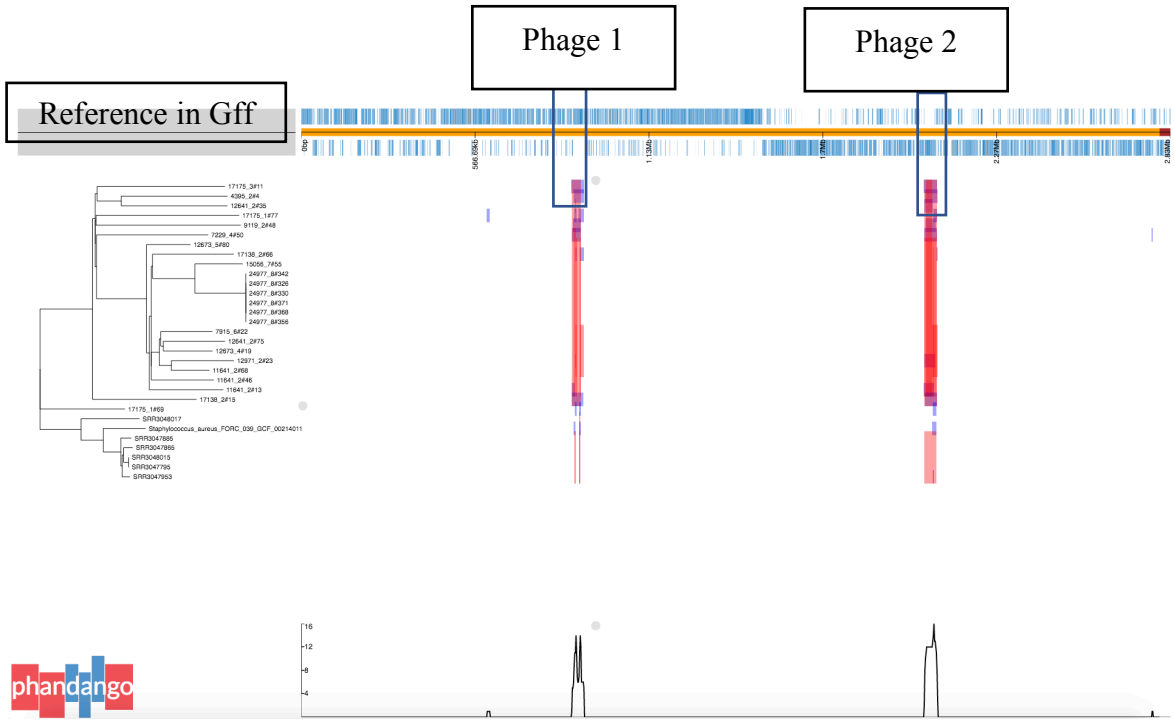


Fig. 14: Phandango visualization of 2 regions of repetitive regions (Red) in pseudogenome alignment of ST188 as identified by Gubbins alongside RAxML phylogenetic tree reconstructed by Gubbins without these regions.

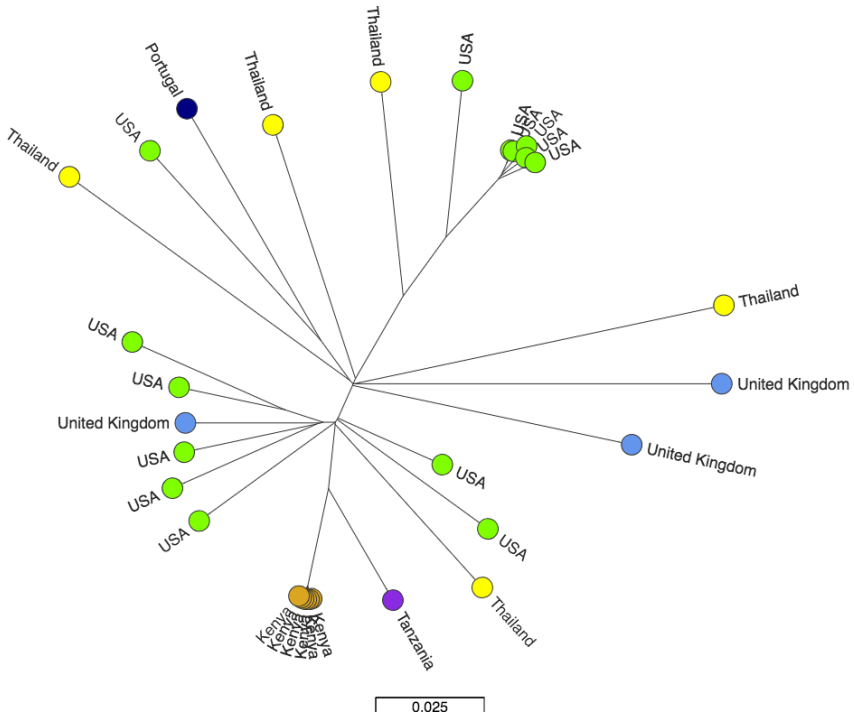


Fig. 15: Midpoint rooted phylogenetic tree of ST188 based on 2552 core genome SNPs on mapping to HongKong draft genome reference. Nodes indicate country of origin.

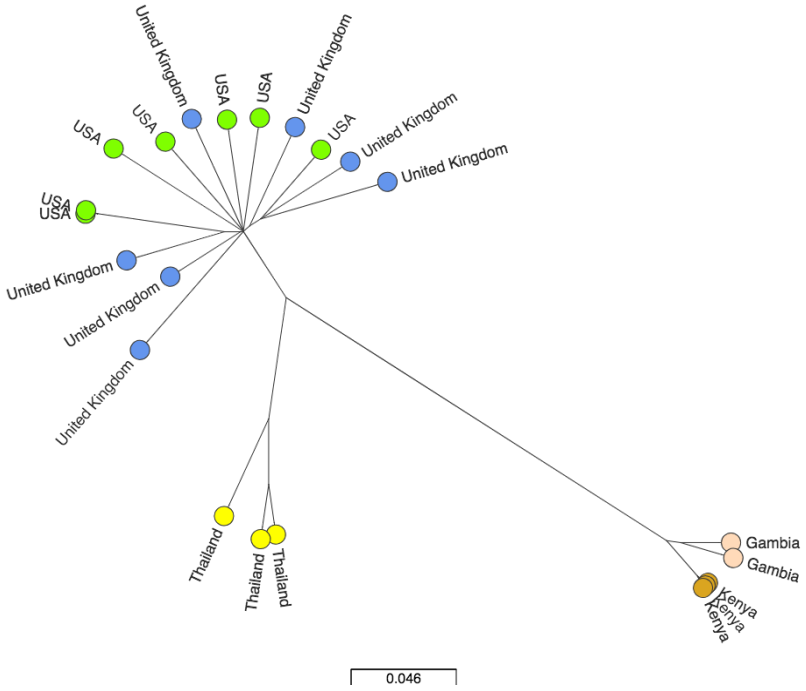


Fig. 16: A Midpoint rooted ML phylogeny of ST25 based on 3011 core genome SNPs on alignment to TW20 reference

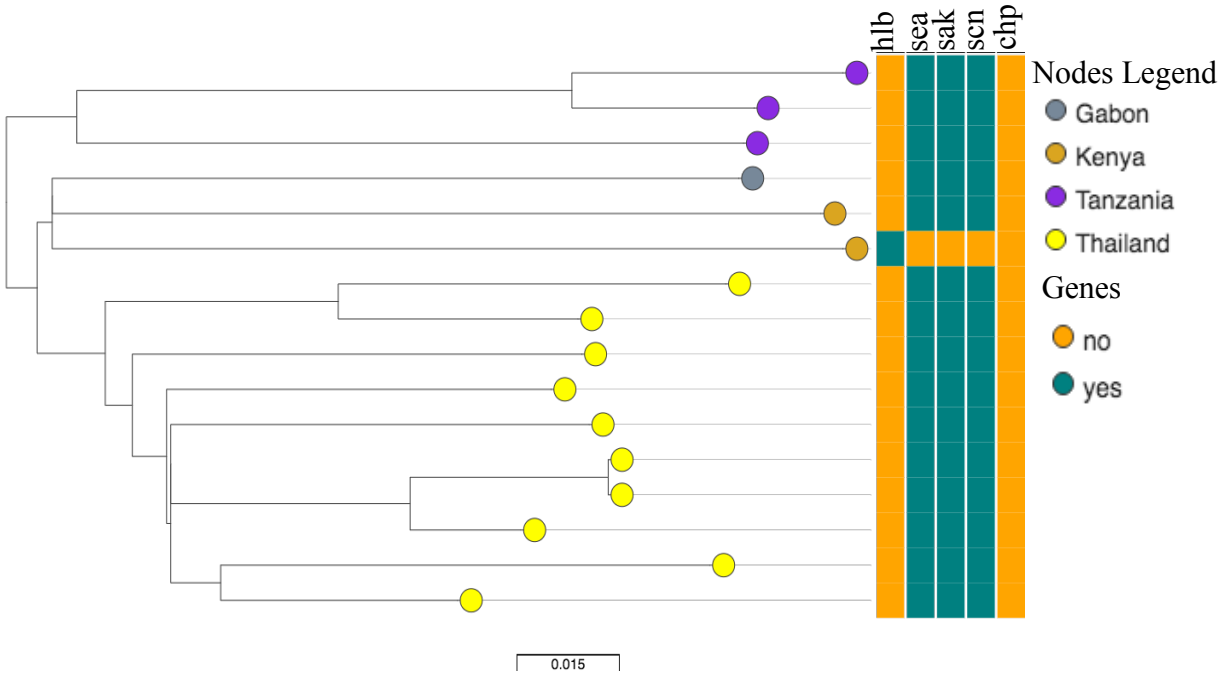


Fig. 17: Midpoint rooted ML phylogenetic tree of ST6 based on 1283 core genome SNPs on alignment to *Staphylococcus aureus* TW20 reference. The tree tips are aligned to presence or absence of integrase prophage group 3 genes. Green color in the heatmap indicate presence and orange color absence

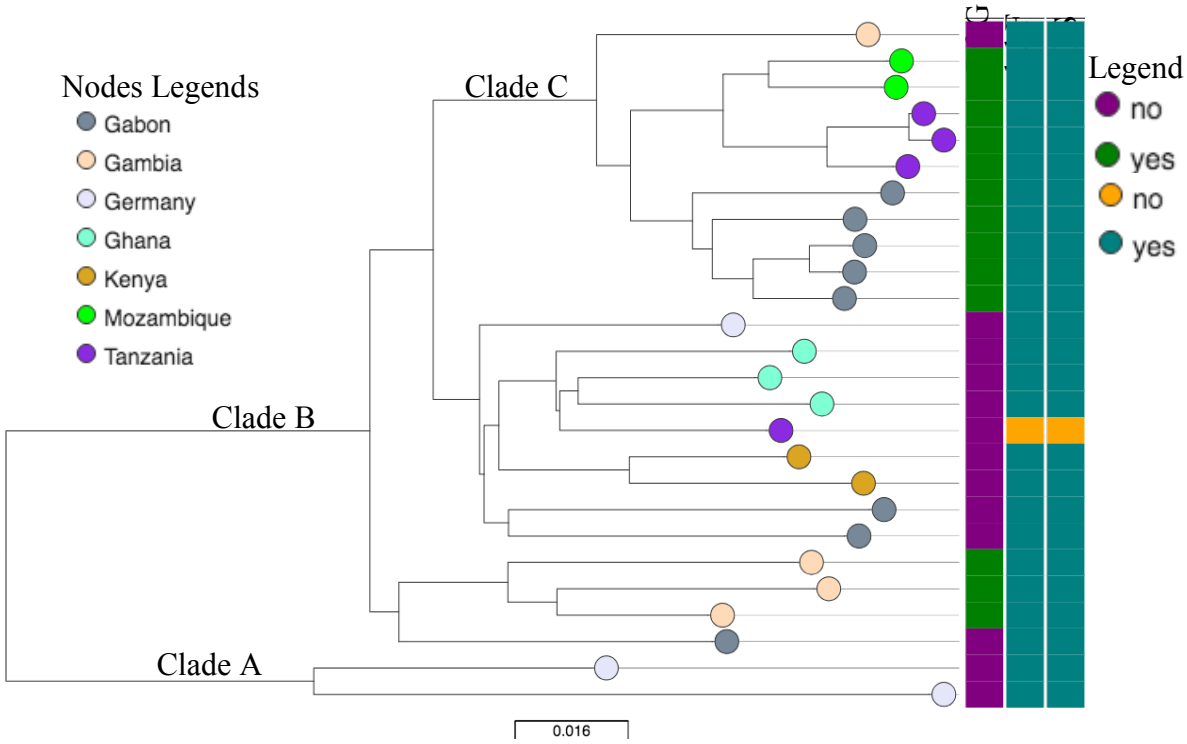


Fig. 18: Midpoint rooted ML phylogenetic tree of ST152 based on 1446 core genome SNPs on alignment to *Staphylococcus aureus* BB155 reference. The tree tips are linked to dfrG and PVL presence and absence genes

Generally, Kenyan isolates were phylogenetically related to other African *S. aureus* strains. For instance, Kenyan isolates of ST188 were closely related to an isolate from Tanzania among 20 genomes collected from six countries representing 4 continents (Fig. 15). A similar observation was demonstrated in ST25 phylogeny (Fig. 16), where three Kenyan and two Gambian strains formed distinct clades among three phylogenetic clades of 22 genomes from across three other continents.

Furthermore, phage type D that are knowns to carry staphylococcal enterotoxin a (*sea*), staphylokinase (*sak*) and staphylococcal complement inhibitory protein (*scn*) (van Wamel et al., 2006b) were present in almost all strains 92% (n=13) of ST6 except one Kenyan isolate that had intact *hlb* although it was sampled from human (Fig. 17) .

Another important observation in the phylogenetic tree of ST152 (Fig. 18) was clade C that had two sub-clusters; one for trimethoprim resistant (*dfrG*) and another for susceptible genomes. The Kenyan strains being trimethoprim susceptible were closely related to Tanzania and Ghana genomes that were also susceptible to the drug. The temperate bacteriophage that harbor *luk*S-PV and *luk*F-PV seem to be conserved in this lineage ST152. The distribution of nucleotide sequences of PVL genes in ST152 were identical with exception of two novel non-synonymous mutations at position 532 and 668 of *lukS-PV*, and position 153 and 789 for *lukF-PV* (Zhao et al., 2016), (Fig. 19) of two Germany strains of clade A (Fig. 18) in the phylogenetic tree.

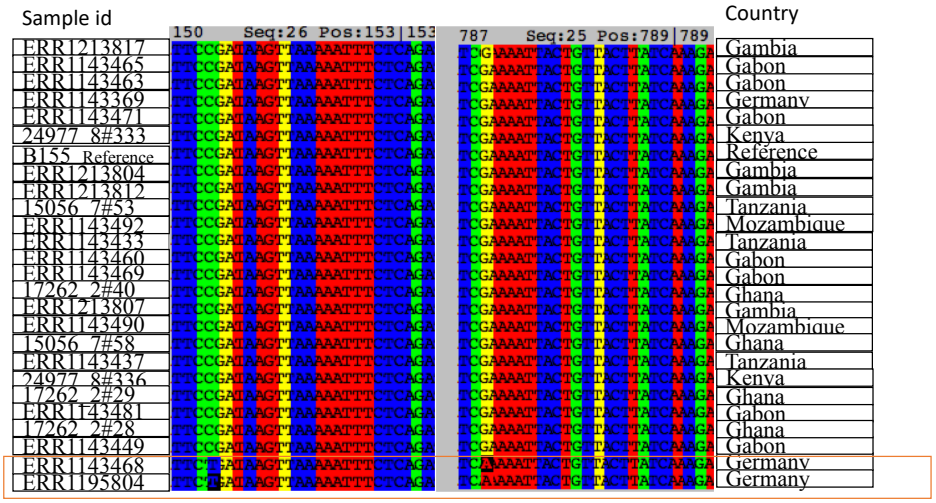


Fig. 19: Multiple sequence alignment of lukF-PV showing non-synonymous SNPs at positions 153 and 789 of two German genomes (bottom highlighted rectangular shape)

3.2.3 Phylogenetic Analyses of ST580 and Livestock-associated ST398 lineage

To find out whether three Kenyan isolates (2 swine and 1 human) of ST580 could be livestock adapted or human adapted lineage, I analyzed their phylogenetic relatedness with 30 global genomes of ST398 that were isolated from 7 countries between 2006 to 2015. The ST580 lineage (ST profile 3-35-48-19-20-26-39) is a double locus variant of ST398 (3-35-19-2-20-26-39) because they differ by only two allelic loci of the seven housekeeping genes used in determination of multi-locus sequence types (MLST) (http://saureus.mlst.net). These 33 strains were mapped to livestock-associated MRSA ST398 reference and subsequently reconstructed phylogenetic based on 8575 core genome SNPs.

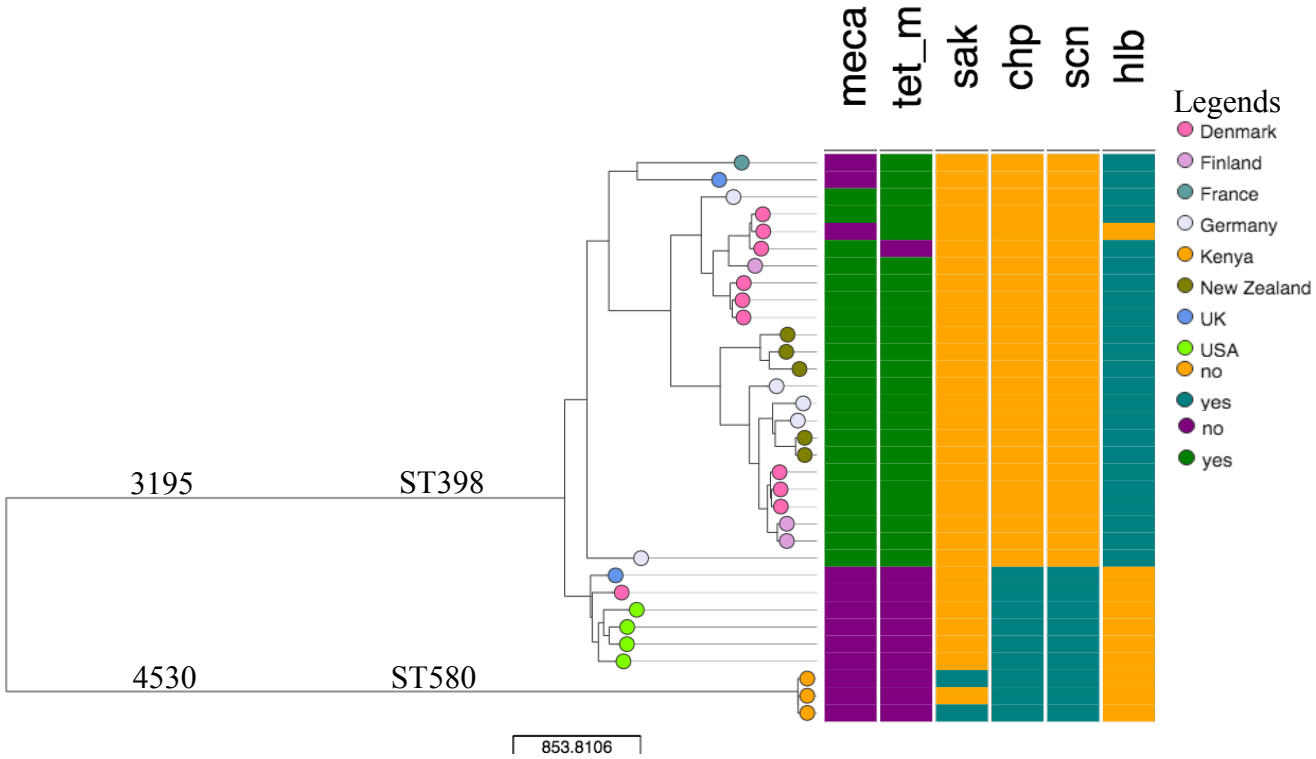


Fig. 20: Maximum Likelihood phylogeny of ST398 and ST580 based on 8575 core genome SNPs that has been optimized with ACCTRAN algorithms. The tree is linked to the heatmap of AMR and IECs genes. The number in the branched show the total number of SNPs in the branch after parsimony.

As expected, the genomes in the phylogeny grouped into 2 distinct clusters (Fig. 20), one coinciding with ST398 and the other with long branch length for 3 strains of ST580. The ST580 Kenyan isolates differ with the reference by an average of 6925 SNPs, suggesting that they are distantly related. Interesting, ST398 global collection strains formed two major clades; the basal clade with strains that were methicillin susceptible and had ß-toxin converting phage type C (positive for *chp* and *scn*), and the top clade with strains that were mostly methicillin (*mec*A) and tetracycline (*tet*M) resistant and had lost phages that carried human immune evasions cluster genes (*scn*, *sak*, *chp*) that are located in ß-hemolysin (*hlb*). This followed a similar observation by Price et al. (Price et al., 2012b) that the human adapted MSSA of ST398 with phages that carried human modulatory proteins, methicillin susceptible, tetracycline susceptible were forming ancestral basal clades. Another striking feature of the phylogenetic tree was that Kenyan ST580 strains were relatively closer (average of 6945 SNPs) to human adapted strains in the basal clade compared to livestock adapted isolates (average of 6970 SNPs differences pairwise) of ST398. This is further supported by strains of ST580 with similar characteristics of isolates of basal clade of ST398 cluster such as methicillin and tetracycline susceptible, negative for *hlb* genes and harboring IECs genes.

3.2.4 Accessory Genomes Analyses of ST580 and ST398

In order to understand further the genetic relatedness of ST580 and ST398, I explored their distribution of genes in their accessory genomes using Roary which bin the genes into core and accessory genomes. The number of genes in the pan-genome was 4084, in which 1910 genes were core genes and 2174 genes were accessory. 2221 genes were discarded that were present in more than 90% of the strains and created a pairwise matrix (supplementary file 7) based on 1863 accessory genes using a custom python script. In order to visualize this output pairwise matrix, R Studio were used to generate heatmap and clustering tree based on the pairwise proportions of number of shared genes in the accessory genomes (fig 21). The darker the color in the heatmap, the higher the number of shared genes between isolates.

The analysis of the accessory genomes inferred by Roary showed that Kenyan isolates of ST580 seemed to share a higher number of accessory genes with human adapted MSSA than livestock adapted MRSA strains of ST398 lineage (Fig. 21). The annotated genes that the

ST580 strains and human-adapted genomes of ST398 shared include autolysin, some groups of transposon-related proteins, hypothetical proteins, lipoproteins, some phages that encode for *chp*, *scn* and *sak* genes, serine aspartate repeat proteins *sdrC*. They differ mostly with livestock adapted genomes in the ST398 lineage in terms of genes that encode for phage proteins, transposase, recombinases proteins, pathogenicity islands proteins, hypothetical proteins among many other mobile genetic elements.

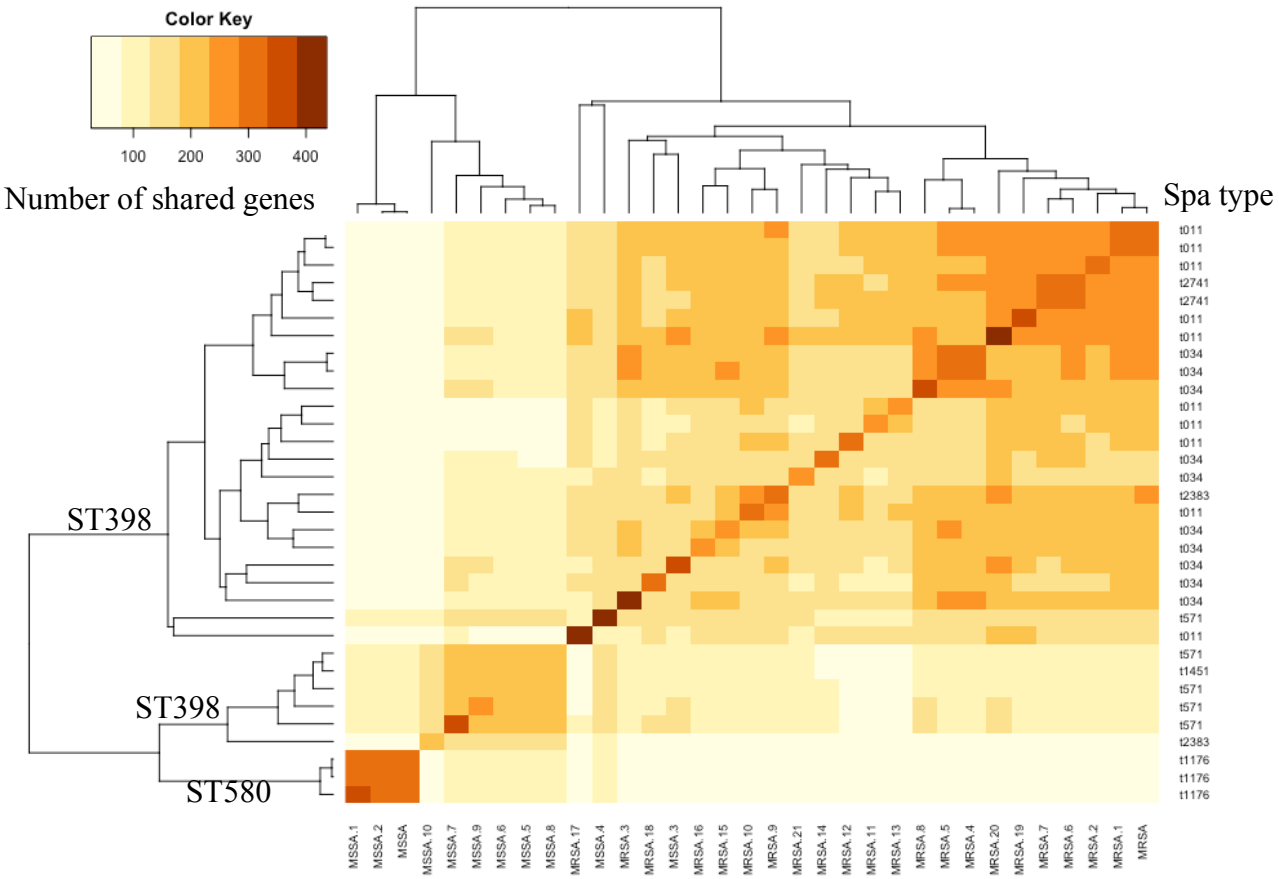


Fig. 21: Heatmap based on the number of shared genes in the accessory genomes. The x-axis labels indicate presence of either MSSA or MRSA while y-axis indicate the spa type of the STs. The color key indicates the proportion of number of shared genes. The genomes were clustered based on the accessory gene distribution

CHAPTER FOUR

DISCUSSION

This is the first study to characterize the whole genome of *Staphylococcus aureus* between humans and pigs in Africa. Moreover, the previous studies that involved characterization of *S. aureus* in Kenya are skewed towards only human clinical isolates and importantly all were carried out using conventional typing tools (Omuse et al., 2016, Aiken et al., 2014). Investigation of Multi-locus sequence type (MLST) and *spa* type revealed a significant degree of genetic diversity of clonal lineages circulating in humans and pigs that were sampled from Kiambu, Kenya between October 2015 and August 2016. Furthermore, analysis showed that humans and swine strains were inter-spread across the phylogeny with some having homogenous genetic polymorphisms. In addition, there was no clear distinctions of antimicrobial resistance and virulence genes between the hosts providing evidence for the need for future genomic surveillance and advancement of one health initiative.

The results of determination of population structure clones provided evidence of a considerable degree of genetic variation of *S. aureus* in Kiambu, Kenya. This was comparable with the recent findings of Omuse et al. (Omuse et al., 2016), who identified 40 *spa* types among 55 clinical strains from Aga Khan University hospital and KEMRI in Nairobi, which are located approximately 30 kilometers (Km) from Kiambu county. These *spa* types included t223, t355 and t091 that were also identified in this study (Fig. 6), suggesting that these strains could represent isolates causing hospital infections in Kenya. Investigation of MLST, identified ST188 as the dominant clone, which was present in both swine and human strains (Fig. 5). This ST188 clone, is frequently isolated in hospitals and community acquired infections in Pan Asian regions (Wang et al., 2018), and has been reported to be of high prevalence among healthy humans and various animals in Africa, including chimpanzee and lemurs in Uganda and Madagascar, respectively (Schaumburg et al., 2013), and goats, cats and cattle in Tunisia (Gharsa et al., 2015) and has been associated with disease outbreak among laboratory mice (Sung et al., 2008, McCarthy et al., 2012). This suggest that this clone has potential to colonize different types of hosts which corresponds with the finding of this study. In addition, the strains of ST188 had been

documented to have low level of antibiotic resistance genes. For instance, none of these African isolates, including Kenyan isolates of this study, were MRSA. However, Wang et al. demonstrated during a follow-up study of this lineage between 2012 and 2014 in China that this clone was gaining antibiotic resistance over time (Wang et al., 2018), hence the need for future continuous surveillance of this lineage.

There was no evidence of antimicrobial resistance genes profiles that were associated with either pigs or humans, as expected considering the small number of the remaining good quality genomes in this study. This is in contrast with recent findings by Richardson et al. where they identified specific classes of antimicrobial resistance genes being significantly associated with some host species suggesting important role of antibiotic selection pressure in humans and animals in evolution and emergence of clones with antimicrobial resistance genes (Richardson et al., 2018). Notably, the high prevalence of *blaZ* gene (Fig. 9) that confers resistance to penicillin and ampicillin was expected in this study because penicillin has been commonly used for treatment of infections in humans and animals since 1940s (Gundogan et al., 2005), and in agreement with previous studies in other African countries (Gitau et al., 2018, Katakweba et al., 2016, Akanbi et al., 2017).

The investigation of virulence determinants identified many important genetic markers (Fig. 10) that had been associated previously with diseases and host adaptations. This include *lukF/S* genes that were found in ST152 and ST15 which have been associated with carriage of PVL genes (Schaumburg et al., 2014b). Furthermore, the ST152 clone is frequently isolated in both community and hospital settings in Sub-Saharan Africa (over 40% of the strains) (Ouedraogo et al., 2016). The analysis of PVL in ST152 lineage showed identical distribution of nucleotides sequences of its encoding genes, *lukF-PV* and *lukS-PV* (Fig. 19) except in two strains of the distant clade (Fig. 18). Although PVL is highly specific for human neutrophils, its detection in a swine isolate is consistent with recent reports in Algeria where the investigators identified 4/19 sheep isolates belonging to ST152 being positive for this gene (Agabou et al., 2017) suggesting that PVL could be conserved in this lineage irrespective of the hosts.

Studies have demonstrated that the type of enterotoxins produced by *S. aureus* could be used to indicate the host origin of the strains. The classical enterotoxins SEA and SEC have been shown to be associated with strains that originated from humans and SED from animals

(Gonzalez et al., 2017, Jones et al., 2002). The identification of swine strains with SEC genes in this study could point to a human origin. The co-existence of newly described enterotoxins *sem, sei, seo sen* and *seg* was unsurprising since they belong to enterotoxin gene clusters (*egc*) that are found in the same genomic islands *vSaß* (Yan et al., 2012). Moreover, loci of these *egc* in the *vSaß* had been shown previously to be highly correlated with the clonal lineages (van Belkum et al., 2006, Chao et al., 2015, Song et al., 2016), and were found in this study in ST25 and ST22 lineages. Furthermore, co-detection of *tsst-1*, causing toxic shock syndrome, with *egc* had also been documented in other previous studies (Song et al., 2016).

Typically, ß-hemolysin converting bacteriophages that harbor immune evasion clusters (IECs) genes *sak*, *scn*, *chp* and *sea* are strongly associated with humans (Verkaik et al., 2011a) and its acquisition or loss had been considered specific genetic markers for host adaptations. The presence of these genes in swine strains of the clonal lineages that were shared with the humans indicate that the isolates are likely to have been transferred recently to swine. Proper surveillance needs to be established to monitor strains carrying these genes since they have the potential to cause infections in humans. In addition, there were two strains with IECs phage type H (associated with livestock adaptations) that belonged to ST6 and ST22 lineages which have been demonstrated in previous studies to colonize animals and get adapted by losing these IECs genes. For instances, in monkeys in Gambia (Senghore et al., 2016), cats and dogs in United Kingdom (Harrison et al., 2014), Guinea pigs, rabbits in Germany (Walther et al., 2008).

Studies in Africa have demonstrated that the major clonal lineages circulating in animals are also present in humans (Lozano et al., 2016a, Schaumburg et al., 2015) and this is also a general observation of this study. In addition to using MLST and *spa* typing to investigate the possible sharing of clones between humans and swine in Kiambu, I also looked for evidence of transmission within and across homesteads based on core genome SNPs. Surprisingly, the strains of swine and humans of rare clones ST789 and ST188, yet to be described as global pandemic lineages for hospitals infections, or livestock associated lineages, were highly similar in terms of SNPs even though some were sampled from different homesteads. This suggests that the strains within each of these clones recently shared a common ancestor and spread across homesteads. Alternatively, the clones could be

stable within the environments. However, a previous study of *S. aureus* strains belonging to ST188 from primates, environment and personnel working in a sanctuary center in the USA reported highly similar polymorphisms of seven strains although the isolates were sampled at different times with different antibiotic resistance profiles (Soge et al., 2016). This could indicate this lineage could be having different lower mutation rates compared to the mutation rates of 1.2×10^{-6} to 2.0×10^{-6} reported for other *S. aureus* lineages (Fitzgerald and Holden, 2016) or 3.3×10^{-6} in ST239 lineage (Harris et al., 2010b).

The livestock-associated ST398 lineage has gained a lot of relevance in public health because of its broad host tropisms, and to cause clinical infections. Therefore, the genetic relatedness of ST398 and its double variant ST580 is of great relevance in epidemiological surveillance in the light of one health concept. To the best of my knowledge, this is the first documentation to use whole genome sequencing data to infer relatedness of these two lineages. The earliest documentation of ST580 in MLST database (http://saureus.mlst.net/), is DCC1185 which was methicillin susceptible and was isolated from human colonization at Lisbon, Portugal in 1997. This current study demonstrated that ST580 (one human and two swine) strains were distantly related to the ST398 reference separated by large number of SNPs with long deep branch in the phylogeny (Fig. 20). Notably, they were closer to the strains of basal human associated clade in ST398 phylogeny suggesting that these strains of ST580 belong to human adapted lineages. Previous studies have associated mobile genetic elements found in the accessory genomes of *S. aureus* with the hosts adaptations (Lindsay, 2010). To further validate this hypothesis, the accessory genomes of strains in ST580 and ST398 lineages were compared. Interestingly, strains of ST580 co-segregated together with isolates of human-adapted ST398 lineage (Fig. 21) based on the distribution of the accessory genes, thus supporting the importance of mobile genetic elements in hosts adaptations.

This study has several limitations. First, the elimination of 71 genomes from the analysis mainly due to contamination impaired the analysis of transmission of *S. aureus* between humans and swine within and across homesteads. Even in the absence of contamination, the selection of only a small number of swine and farmers for sampling in each homestead, and the isolation of a single *S. aureus* colony per sampled subject (thus ignoring within-host diversity) could have resulted in uncertainties in the inference of possible transmission

events between subjects/hosts. Furthermore, lack of information such as sampling date for the isolates, and global positioning system (GPS) location of the homesteads as a result of ethical inconsideration at the time of sampling, could have been useful in suggesting recent transmission events of the strains with similar polymorphisms. Secondly, the cross-sectional design of this study relied on assumptions that the farmers and the pigs were persistent carriers of the pathogen. A longitudinal study of *S. aureus* involving deep sequencing of samples, multiple sampling sites of the host, sampling of other livestock in the farm as well as the environments, and screening of pathogen in the hospitals would have constituted an ideal scenario. However, this is not always possible due to operational and budget constraints. Thirdly, the public genomes that were included in the study were skewed towards the human host and this significantly reduced the power of investigating antibiotic resistance genes that could be associated with a particular host, in addition to examining virulence factors that could be essential in hosts adaptations. Finally, clinical isolates that were sampled in Kiambu county hospital (where most farmers get treatment) between 2013 and 2016 were excluded from the whole-genome sequencing project, and their inclusion could have been of great importance in understanding genetic relatedness of clonal lineages that are present in healthy carriers in the community with those of hospital infections.

CHAPTER FIVE

CONCLUSIONS

To the best of my knowledge, this study is the first to characterize at whole genomic level the strains of *S. aureus* between humans and pigs in Africa. The quality control analysis demonstrated the usefulness of whole genome sequencing (WGS) to identify contaminants from the genomes and mixed infections which could have been missed out using conventional typing methods and other laboratory procedures. The results of *in silico* prediction of *spa* types and MLST revealed higher genetic diversity of *S. aureus* in Kiambu County consistent with previous epidemiological findings in Kenya and Africa in general. Phylogeny analyses showed that some isolates of the same *spa* type and MLST were distantly related and were separated by a large number of SNPs demonstrating usefulness of WGS in providing high accuracy and resolution in epidemiological surveillance studies. The observation that humans and pigs share some of the clones and, in some cases, the genomes share similar polymorphisms, justifies further epidemiological investigations to establish the source and the direction of transmission of *S. aureus* strains. The presence of enterotoxin *sec* and human immune evasions gene clusters (IECs) in majority of the strains, even in swine isolates, suggests these strains were recently acquired from humans and were yet to get adapted in pigs. The analyses of enterotoxin gene cluster (*egc*) and Panton-Valentine leucocidin (PVL) genes suggested that they could be associated with ST25 and ST152 lineages respectively. Investigation of Kenyan isolates with global public genomes showed that they were closely related with strains from other African countries suggesting introductions of strains to and/or from Kenya to other African countries. Despite the small datasets analyzed in this study, it provided the basis for further genomic epidemiological surveillance studies and advancement of one health initiative.

5.1 FUTURE PERSPECTIVES OF THE STUDY

5.1.1 High Resolution and Accuracy of Whole Genome Sequencing (WGS)

The quality control analyses of this study provided evidence of the usefulness of WGS to identify presence and proportions of contaminants in the sequence data. In addition, it displayed mixed infections of *S. aureus* strains in some of the genomes. These factors negatively impacted on initial research questions and aims of the study project. However, these could be avoided in future following established purification protocols, choice of selective media for species identification and working under containment laboratory rooms.

5.1.2 Greater Genetic Diversity of Clonal Lineages

Prediction of 9 sequence types (STs) among only 23 strains suggested greater genetic diversity in terms of clonal lineages circulating in Kiambu region. These STs range from specific geographical region clonal lineages such as ST188 that is commonly isolated in Pan-Asian region to globally pandemic lineages such as ST22, ST97 and ST25. However, these genomes were only from small subsets from the collections of samples that were collected in a particular region of approximately 2500 square kilometers (km^2), as opposed to a wider geographic region of Kenya of about 600,000 km^2. Moreover, I did not include clinical isolates to investigate whether these lineages could extend to cause infections in hospitals. Improving power of statistical significance of clonal lineage diversity require wider representation coverage area of the study which will useful in future.

5.1.3 Sharing of Clonal Lineages Between Pigs and Humans

Even though *spa* type and MLST revealed sharing of clonal lineages between humans and pigs, coupled with inter-spread of swine and human strains across the phylogeny and some with similar polymorphisms, we could not be confident enough to suggest possible transmission events. This is because of limited datasets representing small proportions of the cross-sectional study design. However, designing a longitudinal study that covers wider geographic regions and carrying out deep environmental sampling could be ideal in inferring transmission events and evolution of clones in Kenya. This study demonstrated the existence of more than one *S. aureus* strain in a single host therefore sampling multiple sites and deep

sequencing of the colonies of each host are warranted in the future to understand within host diversity. I also realized that epidemiological information of the host such as GPS locations of the farm and the date of isolation, that was lacking in this study could be important in answering transmission questions and estimating the mutation rates of some of the lineages.

5.1.4 Prediction of Antimicrobial Resistance Genes using WGS

Previous studies have demonstrated the usefulness of whole genome sequencing for the prediction of presence of antimicrobial resistance (AMR) genes and the phenotypic expression in routine laboratory antibiotic susceptibility testing (AST) of *S. aureus* strains (Gordon et al., 2014, Aanensen et al., 2016a). Their findings showed WGS could in future replace laborious procedures involved in AST determination in the laboratory although may not be feasible in low- and middle-income countries as of now because of limitations such as expertise knowledge and cost that accompany WGS. Increasing the number of strains sequenced as well as ensuring that the same colonies are used for both AST and DNA extraction and sequencing will sufficiently increase the power of the concordance analysis between *in silico* prediction of AMR and AST. Furthermore, it will be useful in reducing knowledge gaps of the effect of SNPs in AMR genes especially from *S. aureus* strains collected from tropical climatic environments as well as understanding the associations of certain AMR genes profiles with a particular host.

5.1.5 Genetic Relatedness of Kenyan Isolates with Strains of other African Countries

Phylogenetic analyses of human and swine strains of this study with global publicly genomes revealed close genetic relatedness of Kenyan isolates with genomes of other African countries. It will be interesting in the future to carry out epidemiological studies of *S. aureus* in Sub-Saharan Africa that involve large scale genomic analyses of strains sampled longitudinally from diverse countries. This will help in establishing the directions of movement of *S. aureus* strains between countries as well as determining clonal lineages that could be dominant in some regions and temporal signals investigation of certain genes located in mobile genetic elements for adaptations of those dominant lineages.

REFERENCES:

AANENSEN, D. M., FEIL, E. J., HOLDEN, M. T., DORDEL, J., YEATS, C. A., FEDOSEJEV, A., GOATER, R., CASTILLO-RAMIREZ, S., CORANDER, J., COLIJN, C., CHLEBOWICZ, M. A., SCHOULS, L., HECK, M., PLUISTER, G., RUIMY, R., KAHLMETER, G., AHMAN, J., MATUSCHEK, E., FRIEDRICH, A. W., PARKHILL, J., BENTLEY, S. D., SPRATT, B. G. & GRUNDMANN, H. 2016a. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. *MBio,* 7.

ABDEL-MOEIN, K. A., EL-HARIRI, M. & SAMIR, A. 2012. Methicillin-resistant Staphylococcus aureus: an emerging pathogen of pets in Egypt with a public health burden. *Transbound Emerg Dis,* 59**,** 331-5.

ABDULGADER, S. M., SHITTU, A. O., NICOL, M. P. & KABA, M. 2015. Molecular epidemiology of Methicillin-resistant Staphylococcus aureus in Africa: a systematic review. *Front Microbiol,* 6**,** 348.

ACTON, D. S., PLAT-SINNIGE, M. J., VAN WAMEL, W., DE GROOT, N. & VAN BELKUM, A. 2009. Intestinal carriage of Staphylococcus aureus: how does its frequency compare with that of nasal carriage and what is its clinical impact? *Eur J Clin Microbiol Infect Dis,* 28**,** 115-27.

ADEGOKE, A. A. & OKOH, A. I. 2014. Species diversity and antibiotic resistance properties of Staphylococcus of farm animal origin in Nkonkobe Municipality, South Africa. *Folia Microbiol (Praha),* 59**,** 133-40.

AGABOU, A., OUCHENANE, Z., NGBA ESSEBE, C., KHEMISSI, S., CHEHBOUB, M. T. E., CHEHBOUB, I. B., SOTTO, A., DUNYACH-REMY, C. & LAVIGNE, J.-P. 2017. Emergence of Nasal Carriage of ST80 and ST152 PVL+ Staphylococcus aureus Isolates from Livestock in Algeria. *Toxins,* 9**,** 303.

AIKEN, A. M., MUTUKU, I. M., SABAT, A. J., AKKERBOOM, V., MWANGI, J., SCOTT, J. A. G., MORPETH, S. C., FRIEDRICH, A. W. & GRUNDMANN, H. 2014. Carriage of Staphylococcus aureus in Thika Level 5 Hospital, Kenya: a cross-sectional study. *Antimicrobial Resistance and Infection Control,* 3**,** 22.

AKANBI, O. E., NJOM, H. A., FRI, J., OTIGBU, A. C. & CLARKE, A. M. 2017. Antimicrobial Susceptibility of Staphylococcus aureus Isolated from Recreational Waters and Beach Sand in Eastern Cape Province of South Africa. *International Journal of Environmental Research and Public Health,* 14**,** 1001.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ARGIMON, S., ABUDAHAB, K., GOATER, R. J., FEDOSEJEV, A., BHAI, J., GLASNER, C., FEIL, E. J., HOLDEN, M. T., YEATS, C. A., GRUNDMANN, H., SPRATT, B. G. & AANENSEN, D. M. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom,* 2**,** e000093.

ARGUDÍN, M. Á., MENDOZA, M. C. & RODICIO, M. R. 2010. Food Poisoning and Staphylococcus aureus Enterotoxins. *Toxins,* 2**,** 1751-1773.

ARMAND-LEFEVRE, L., RUIMY, R. & ANDREMONT, A. 2005. Clonal comparison of Staphylococcus aureus isolates from healthy pig farmers, human controls, and pigs. *Emerg Infect Dis,* 11**,** 711-4.

BABA, T., TAKEUCHI, F., KURODA, M., YUZAWA, H., AOKI, K., OGUCHI, A., NAGAI, Y., IWAMA, N., ASANO, K., NAIMI, T., KURODA, H., CUI, L., YAMAMOTO, K. & HIRAMATSU, K. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet,* 359**,** 1819-27.

BALLHAUSEN, B., KRIEGESKORTE, A., VAN ALEN, S., JUNG, P., KÖCK, R., PETERS, G., BISCHOFF, M. & BECKER, K. 2017. The pathogenicity and host adaptation of livestock-associated MRSA CC398. *Veterinary Microbiology,* 200**,** 39-45.

BANNERMAN, T. L., HANCOCK, G. A., TENOVER, F. C. & MILLER, J. M. 1995. Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of Staphylococcus aureus. *J Clin Microbiol,* 33**,** 551-5.

BEN ZAKOUR, N. L., GUINANE, C. M. & FITZGERALD, J. R. 2008. Pathogenomics of the staphylococci: insights into niche adaptation and the emergence of new virulent strains. *FEMS Microbiol Lett,* 289**,** 1-12.

BERGSTROM, K., BENGTSSON, B., NYMAN, A. & GRONLUND ANDERSSON, U. 2013. Longitudinal study of horses for carriage of methicillin-resistant Staphylococcus aureus following wound infections. *Vet Microbiol,* 163**,** 388-91.

BERKLEY, J. A., BEJON, P., MWANGI, T., GWER, S., MAITLAND, K., WILLIAMS, T. N., MOHAMMED, S., OSIER, F., KINYANJUI, S., FEGAN, G., LOWE, B. S., ENGLISH, M., PESHU, N., MARSH, K. & NEWTON, C. R. 2009. HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria. *Clin Infect Dis,* 49**,** 336-43.

BLAIR, J. M., WEBBER, M. A., BAYLAY, A. J., OGBOLU, D. O. & PIDDOCK, L. J. 2015. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol,* 13**,** 42-51.

BLETZ, S., MELLMANN, A., ROTHGANGER, J. & HARMSEN, D. 2015. Ensuring backwards compatibility: traditional genotyping efforts in the era of whole genome sequencing. *Clin Microbiol Infect,* 21**,** 347.e1-4.

BOETZER, M., HENKEL, C. V., JANSEN, H. J., BUTLER, D. & PIROVANO, W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27**,** 578-9.

BOETZER, M. & PIROVANO, W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol,* 13**,** R56.

BOYLE-VAVRA, S. & DAUM, R. S. 2006. Community-acquired methicillin-resistant Staphylococcus aureus: the role of Panton–Valentine leukocidin. *Laboratory Investigation,* 87**,** 3.

BRADLEY, A. 2002. Bovine mastitis: an evolving disease. *Vet J,* 164**,** 116-28.

BRIGIDO MDE, M., BARARDI, C. R., BONJARDIN, C. A., SANTOS, C. L., JUNQUEIRA, M. L. & BRENTANI, R. R. 1991. Nucleotide sequence of a variant protein A of Staphylococcus aureus suggests molecular heterogeneity among strains. *J Basic Microbiol,* 31**,** 337-45.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

CHAIRAT, S., GHARSA, H., LOZANO, C., GOMEZ SANZ, E., GÓMEZ, P., ZARAZAGA, M., BOUDABOUS, A., TORRES, C. & KARIM, B. S. 2015.

Characterization of Staphylococcus aureus from Raw Meat Samples in Tunisia: Detection of Clonal Lineage ST398 from the African Continent.

CHAMBERS, H. F. & DELEO, F. R. 2009. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol,* 7**,** 629-41.

CHANG, S., SIEVERT, D. M., HAGEMAN, J. C., BOULTON, M. L., TENOVER, F. C., DOWNES, F. P., SHAH, S., RUDRIK, J. T., PUPP, G. R., BROWN, W. J., CARDO, D. & FRIDKIN, S. K. 2003. Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. *N Engl J Med,* 348**,** 1342-7.

CHAO, G., BAO, G., CAO, Y., YAN, W., WANG, Y., ZHANG, X., ZHOU, L. & WU, Y. 2015. Prevalence and diversity of enterotoxin genes with genetic background of Staphylococcus aureus isolates from different origins in China. *International Journal of Food Microbiology,* 211**,** 142-147.

CHRISTIANE, G., CHRISTIANE, W., URSULA, F. & CHRISTIANE, W. 2006. Extensive phage dynamics in Staphylococcus aureus contributes to adaptation to the human host during infection. *Molecular Microbiology,* 61**,** 1673-1685.

CHUANG, Y. Y. & HUANG, Y. C. 2015. Livestock-associated meticillin-resistant Staphylococcus aureus in Asia: an emerging issue? *Int J Antimicrob Agents,* 45**,** 334-40.

CLAEYS, K. C., LAGNF, A. M., HALLESY, J. A., COMPTON, M. T., GRAVELIN, A. L., DAVIS, S. L. & RYBAK, M. J. 2016. Pneumonia Caused by Methicillin-Resistant Staphylococcus aureus: Does Vancomycin Heteroresistance Matter? *Antimicrob Agents Chemother,* 60**,** 1708-16.

COLL, F., HARRISON, E. M., TOLEMAN, M. S., REUTER, S., RAVEN, K. E., BLANE, B., PALMER, B., KAPPELER, A. R. M., BROWN, N. M., TÖRÖK, M. E., PARKHILL, J. & PEACOCK, S. J. 2017. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Science translational medicine,* 9**,** eaak9745.

COOKSON, B. D., APARICIO, P., DEPLANO, A., STRUELENS, M., GOERING, R. & MARPLES, R. 1996. Inter-centre comparison of pulsed-field gel electrophoresis for the typing of methicillin-resistant Staphylococcus aureus. *J Med Microbiol,* 44**,** 179-84.

CROUCHER, N. J., PAGE, A. J., CONNOR, T. R., DELANEY, A. J., KEANE, J. A., BENTLEY, S. D., PARKHILL, J. & HARRIS, S. R. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research,* 43**,** e15-e15.

CUNY, C., ABDELBARY, M., LAYER, F., WERNER, G. & WITTE, W. 2015. Prevalence of the immune evasion gene cluster in Staphylococcus aureus CC398. *Vet Microbiol,* 177**,** 219-23.

DECLERCQ, P., PETRÉ, D., GORDTS, B. & VOSS, A. 2008. Complicated Community-Acquired Soft Tissue Infection by MRSA from Porcine Origin. *Infection,* 36**,** 590-592.

DEVRIESE, L. A., VAN DAMME, L. R. & FAMEREE, L. 1972. Methicillin (cloxacillin)-resistant Staphylococcus aureus strains isolated from bovine mastitis cases. *Zentralbl Veterinarmed B,* 19**,** 598-605.

DINGES, M. M., ORWIN, P. M. & SCHLIEVERT, P. M. 2000. Exotoxins of Staphylococcus aureus. *Clin Microbiol Rev,* 13**,** 16-34, table of contents.

EDSLEV, S. M., CLAUSEN, M.-L., AGNER, T., STEGGER, M. & ANDERSEN, P. S. 2018. Genomic analysis reveals different mechanisms of fusidic acid resistance in Staphylococcus aureus from Danish atopic dermatitis patients. *Journal of Antimicrobial Chemotherapy,* 73**,** 856-861.

ENRIGHT, M. C., DAY, N. P., DAVIES, C. E., PEACOCK, S. J. & SPRATT, B. G. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *J Clin Microbiol,* 38**,** 1008-15.

FALAGAS, M. E., KARAGEORGOPOULOS, D. E., LEPTIDIS, J. & KORBILA, I. P. 2013. MRSA in Africa: filling the global map of antimicrobial resistance. *PLoS One,* 8**,** e68024.

FALL, C., SECK, A., RICHARD, V., NDOUR, M., SEMBENE, M., LAURENT, F. & BREUREC, S. 2012. Epidemiology of Staphylococcus aureus in pigs and farmers in the largest farm in Dakar, Senegal. *Foodborne Pathog Dis,* 9**,** 962-5.

FEIL, E. J., COOPER, J. E., GRUNDMANN, H., ROBINSON, D. A., ENRIGHT, M. C., BERENDT, T., PEACOCK, S. J., SMITH, J. M., MURPHY, M., SPRATT, B. G.,

MOORE, C. E. & DAY, N. P. 2003. How clonal is Staphylococcus aureus? *J Bacteriol,* 185**,** 3307-16.

FEIL, E. J., LI, B. C., AANENSEN, D. M., HANAGE, W. P. & SPRATT, B. G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol,* 186**,** 1518-30.

FISHER, E. L., OTTO, M. & CHEUNG, G. Y. C. 2018. Basis of Virulence in Enterotoxin-Mediated Staphylococcal Food Poisoning. *Frontiers in Microbiology,* 9**,** 436.

FITZGERALD, J. R. 2012. Livestock-associated Staphylococcus aureus: origin, evolution and public health threat. *Trends Microbiol,* 20**,** 192-8.

FITZGERALD, J. R. & HOLDEN, M. T. 2016. Genomics of Natural Populations of Staphylococcus aureus. *Annu Rev Microbiol,* 70**,** 459-78.

FLOYD, J. L., SMITH, K. P., KUMAR, S. H., FLOYD, J. T. & VARELA, M. F. 2010. LmrS is a multidrug efflux pump of the major facilitator superfamily from Staphylococcus aureus. *Antimicrob Agents Chemother,* 54**,** 5406-12.

FOSTER, T. J., GEOGHEGAN, J. A., GANESH, V. K. & HÖÖK, M. 2014. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. *Nature reviews. Microbiology,* 12**,** 49-62.

FOSTER, T. J. & HOOK, M. 1998. Surface protein adhesins of Staphylococcus aureus. *Trends Microbiol,* 6**,** 484-8.

FU, L., NIU, B., ZHU, Z., WU, S. & LI, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics,* 28**,** 3150-2.

GARCIA-ALVAREZ, L., HOLDEN, M. T., LINDSAY, H., WEBB, C. R., BROWN, D. F., CURRAN, M. D., WALPOLE, E., BROOKS, K., PICKARD, D. J., TEALE, C., PARKHILL, J., BENTLEY, S. D., EDWARDS, G. F., GIRVAN, E. K., KEARNS, A. M., PICHON, B., HILL, R. L., LARSEN, A. R., SKOV, R. L., PEACOCK, S. J., MASKELL, D. J. & HOLMES, M. A. 2011. Meticillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect Dis,* 11**,** 595-603.

GEBREYES, W. A., DUPOUY-CAMET, J., NEWPORT, M. J., OLIVEIRA, C. J., SCHLESINGER, L. S., SAIF, Y. M., KARIUKI, S., SAIF, L. J., SAVILLE, W.,

WITTUM, T., HOET, A., QUESSY, S., KAZWALA, R., TEKOLA, B., SHRYOCK, T., BISESI, M., PATCHANEE, P., BOONMAR, S. & KING, L. J. 2014. The global one health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings. *PLoS Negl Trop Dis,* 8**,** e3257.

GHARSA, H., BEN SLAMA, K., GOMEZ-SANZ, E., LOZANO, C., ZARAZAGA, M., MESSADI, L., BOUDABOUS, A. & TORRES, C. 2015. Molecular characterization of Staphylococcus aureus from nasal samples of healthy farm animals and pets in Tunisia. *Vector Borne Zoonotic Dis,* 15**,** 109-15.

GILLET, Y., ISSARTEL, B., VANHEMS, P., FOURNET, J. C., LINA, G., BES, M., VANDENESCH, F., PIEMONT, Y., BROUSSE, N., FLORET, D. & ETIENNE, J. 2002. Association between Staphylococcus aureus strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet,* 359**,** 753-9.

GITAU, W., MASIKA, M., MUSYOKI, M., MUSEVE, B. & MUTWIRI, T. 2018. Antimicrobial susceptibility pattern of Staphylococcus aureus isolates from clinical specimens at Kenyatta National Hospital. *BMC Res Notes,* 11**,** 226.

GONCALVES DA SILVA, A., BAINES, S. L., CARTER, G. P., HEFFERNAN, H., FRENCH, N. P., REN, X., SEEMANN, T., BULACH, D., KWONG, J., STINEAR, T. P., HOWDEN, B. P. & WILLIAMSON, D. A. 2017. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant Staphylococcus aureus. *Microb Genom,* 3**,** e000105.

GONZALEZ, A. G. M., MARQUES, L. M. P., GOMES, M., BELTRAO, J., PINHEIRO, M. G., ESPER, L. M. R., PAULA, G. R., TEIXEIRA, L. A. & AGUIAR-ALVES, F. 2017. Methicillin-resistant Staphylococcus aureus in minas frescal cheese: evaluation of classic enterotoxin genes, antimicrobial resistance and clonal diversity. *FEMS Microbiol Lett,* 364.

GORDON, N. C., PRICE, J. R., COLE, K., EVERITT, R., MORGAN, M., FINNEY, J., KEARNS, A. M., PICHON, B., YOUNG, B., WILSON, D. J., LLEWELYN, M. J., PAUL, J., PETO, T. E. A., CROOK, D. W., WALKER, A. S. & GOLUBCHIK, T.

2014. Prediction of Staphylococcus aureus Antimicrobial Resistance by Whole-Genome Sequencing. *Journal of Clinical Microbiology,* 52**,** 1182-1191.

GRAVE, K., TORREN-EDO, J. & MACKAY, D. 2010. Comparison of the sales of veterinary antibacterial agents between 10 European countries. *J Antimicrob Chemother,* 65**,** 2037-40.

GRAVELAND, H., WAGENAAR, J. A., BERGS, K., HEESTERBEEK, H. & HEEDERIK, D. 2011. Persistence of livestock associated MRSA CC398 in humans is dependent on intensity of animal contact. *PLoS One,* 6**,** e16830.

GUNDOGAN, N., CITAK, S., YUCEL, N. & DEVREN, A. 2005. A note on the incidence and antibiotic resistance of Staphylococcus aureus isolated from meat and chicken samples. *Meat Science,* 69**,** 807-810.

GUPTA, S. K., PADMANABHAN, B. R., DIENE, S. M., LOPEZ-ROJAS, R., KEMPF, M., LANDRAUD, L. & ROLAIN, J. M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother,* 58**,** 212-20.

GUSS, B., UHLEN, M., NILSSON, B., LINDBERG, M., SJOQUIST, J. & SJODAHL, J. 1984. Region X, the cell-wall-attachment part of staphylococcal protein A. *Eur J Biochem,* 138**,** 413-20.

HADFIELD, J., CROUCHER, N. J., GOATER, R. J., ABUDAHAB, K., AANENSEN, D. M. & HARRIS, S. R. 2018. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics,* 34**,** 292-293.

HALLIN, M., FRIEDRICH, A. W. & STRUELENS, M. J. 2009. spa typing for epidemiological surveillance of Staphylococcus aureus. *Methods Mol Biol,* 551**,** 189-202.

HARRIS, S. R., FEIL, E. J., HOLDEN, M. T., QUAIL, M. A., NICKERSON, E. K., CHANTRATITA, N., GARDETE, S., TAVARES, A., DAY, N., LINDSAY, J. A., EDGEWORTH, J. D., DE LENCASTRE, H., PARKHILL, J., PEACOCK, S. J. & BENTLEY, S. D. 2010a. Evolution of MRSA during hospital transmission and intercontinental spread. *Science,* 327**,** 469-74.

HARRIS, S. R., FEIL, E. J., HOLDEN, M. T. G., QUAIL, M. A., NICKERSON, E. K., CHANTRATITA, N., GARDETE, S., TAVARES, A., DAY, N., LINDSAY, J. A.,

EDGEWORTH, J. D., DE LENCASTRE, H., PARKHILL, J., PEACOCK, S. J. & BENTLEY, S. D. 2010b. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science (New York, N.Y.),* 327**,** 469-474.

HARRISON, E. M., WEINERT, L. A., HOLDEN, M. T., WELCH, J. J., WILSON, K., MORGAN, F. J., HARRIS, S. R., LOEFFLER, A., BOAG, A. K., PEACOCK, S. J., PATERSON, G. K., WALLER, A. S., PARKHILL, J. & HOLMES, M. A. 2014. A shared population of epidemic methicillin-resistant Staphylococcus aureus 15 circulates in humans and companion animals. *MBio,* 5**,** e00985-13.

HENNEKINNE, J.-A., DE BUYSER, M.-L. & DRAGACCI, S. 2012. Staphylococcus aureus and its food poisoning toxins: characterization and outbreak investigation. *FEMS Microbiology Reviews,* 36**,** 815-836.

HERRMANN, M., ABDULLAH, S., ALABI, A., ALONSO, P., FRIEDRICH, A. W., FUHR, G., GERMANN, A., KERN, W. V., KREMSNER, P. G., MANDOMANDO, I., MELLMANN, A. C., PLUSCHKE, G., RIEG, S., RUFFING, U., SCHAUMBURG, F., TANNER, M., PETERS, G., VON BRIESEN, H., VON EIFF, C., VON MULLER, L. & GROBUSCH, M. P. 2013. Staphylococcal disease in Africa: another neglected 'tropical' disease. *Future Microbiol,* 8**,** 17-26.

HERSCHLEB, J., ANANIEV, G. & SCHWARTZ, D. C. 2007. Pulsed-field gel electrophoresis. *Nat Protoc,* 2**,** 677-84.

HEYSELL, S. K., SHENOI, S. V., CATTERICK, K., THOMAS, T. A. & FRIEDLAND, G. 2011. Prevalence of methicillin-resistant Staphylococcus aureus nasal carriage among hospitalised patients with tuberculosis in rural Kwazulu-Natal. *S Afr Med J,* 101**,** 332-4.

HOLDEN, M. T., FEIL, E. J., LINDSAY, J. A., PEACOCK, S. J., DAY, N. P., ENRIGHT, M. C., FOSTER, T. J., MOORE, C. E., HURST, L., ATKIN, R., BARRON, A., BASON, N., BENTLEY, S. D., CHILLINGWORTH, C., CHILLINGWORTH, T., CHURCHER, C., CLARK, L., CORTON, C., CRONIN, A., DOGGETT, J., DOWD, L., FELTWELL, T., HANCE, Z., HARRIS, B., HAUSER, H., HOLROYD, S., JAGELS, K., JAMES, K. D., LENNARD, N., LINE, A., MAYES, R., MOULE, S., MUNGALL, K., ORMOND, D., QUAIL, M. A., RABBINOWITSCH, E., RUTHERFORD, K., SANDERS, M., SHARP, S., SIMMONDS, M., STEVENS, K.,

WHITEHEAD, S., BARRELL, B. G., SPRATT, B. G. & PARKHILL, J. 2004a. Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A,* 101**,** 9786-91.

HOLDEN, M. T., HSU, L. Y., KURT, K., WEINERT, L. A., MATHER, A. E., HARRIS, S. R., STROMMENGER, B., LAYER, F., WITTE, W., DE LENCASTRE, H., SKOV, R., WESTH, H., ZEMLICKOVA, H., COOMBS, G., KEARNS, A. M., HILL, R. L., EDGEWORTH, J., GOULD, I., GANT, V., COOKE, J., EDWARDS, G. F., MCADAM, P. R., TEMPLETON, K. E., MCCANN, A., ZHOU, Z., CASTILLO-RAMIREZ, S., FEIL, E. J., HUDSON, L. O., ENRIGHT, M. C., BALLOUX, F., AANENSEN, D. M., SPRATT, B. G., FITZGERALD, J. R., PARKHILL, J., ACHTMAN, M., BENTLEY, S. D. & NUBEL, U. 2013a. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Res,* 23**,** 653-64.

HOLDEN, M. T. G., FEIL, E. J., LINDSAY, J. A., PEACOCK, S. J., DAY, N. P. J., ENRIGHT, M. C., FOSTER, T. J., MOORE, C. E., HURST, L., ATKIN, R., BARRON, A., BASON, N., BENTLEY, S. D., CHILLINGWORTH, C., CHILLINGWORTH, T., CHURCHER, C., CLARK, L., CORTON, C., CRONIN, A., DOGGETT, J., DOWD, L., FELTWELL, T., HANCE, Z., HARRIS, B., HAUSER, H., HOLROYD, S., JAGELS, K., JAMES, K. D., LENNARD, N., LINE, A., MAYES, R., MOULE, S., MUNGALL, K., ORMOND, D., QUAIL, M. A., RABBINOWITSCH, E., RUTHERFORD, K., SANDERS, M., SHARP, S., SIMMONDS, M., STEVENS, K., WHITEHEAD, S., BARRELL, B. G., SPRATT, B. G. & PARKHILL, J. 2004b. Complete genomes of two clinical Staphylococcus aureus strains: Evidence for the rapid evolution of virulence and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 9786-9791.

HOLDEN, M. T. G., HSU, L.-Y., KURT, K., WEINERT, L. A., MATHER, A. E., HARRIS, S. R., STROMMENGER, B., LAYER, F., WITTE, W., DE LENCASTRE, H., SKOV, R., WESTH, H., ZEMLICKOVÁ, H., COOMBS, G., KEARNS, A. M., HILL, R. L. R., EDGEWORTH, J., GOULD, I., GANT, V., COOKE, J.,

EDWARDS, G. F., MCADAM, P. R., TEMPLETON, K. E., MCCANN, A., ZHOU, Z., CASTILLO-RAMÍREZ, S., FEIL, E. J., HUDSON, L. O., ENRIGHT, M. C., BALLOUX, F., AANENSEN, D. M., SPRATT, B. G., FITZGERALD, J. R., PARKHILL, J., ACHTMAN, M., BENTLEY, S. D. & NÜBEL, U. 2013b. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome research,* 23**,** 653-664.

HOLMES, M. A. & ZADOKS, R. N. 2011. Methicillin resistant S. aureus in human and bovine mastitis. *J Mammary Gland Biol Neoplasia,* 16**,** 373-82.

HUANG, Y., NIU, B., GAO, Y., FU, L. & LI, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics,* 26**,** 680-2.

HUMPHREYS, H. 2012. 15 - Staphylococcus: Skin infections; osteomyelitis; bloodstream infection; food poisoning; foreign body infections; MRSA. *In:* GREENWOOD, D., BARER, M., SLACK, R. & IRVING, W. (eds.) *Medical Microbiology (Eighteenth Edition).* Edinburgh: Churchill Livingstone.

HUNT, M., MATHER, A. E., SANCHEZ-BUSO, L., PAGE, A. J., PARKHILL, J., KEANE, J. A. & HARRIS, S. R. 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom,* 3**,** e000131.

HYATT, D., CHEN, G. L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics,* 11**,** 119.

IKAWATY, R., BROUWER, E., DUIJKEREN, E., MEVIUS, D., VERHOEF, J. & FLUIT, A. C. 2010. *Virulence Factors of Genotyped Bovine Mastitis Staphylococcus aureus Isolates in The Netherlands*.

J M PATTI, B L ALLEN, M J MCGAVIN, A. & HOOK, M. 1994. MSCRAMM-Mediated Adherence of Microorganisms to Host Tissues. *Annual Review of Microbiology,* 48**,** 585-617.

JAMROZY, D., COLL, F., MATHER, A. E., HARRIS, S. R., HARRISON, E. M., MACGOWAN, A., KARAS, A., ELSTON, T., ESTEE TOROK, M., PARKHILL, J. & PEACOCK, S. J. 2017. Evolution of mobile genetic element composition in an epidemic methicillin-resistant Staphylococcus aureus: temporal changes correlated with frequent loss and gain events. *BMC Genomics,* 18**,** 684.

JEVONS, M. P. 1961. “Celbenin” - resistant Staphylococci. *British Medical Journal,* 1**,** 124-125.

JONES, T. F., KELLUM, M. E., PORTER, S. S., BELL, M. & SCHAFFNER, W. 2002. An outbreak of community-acquired foodborne illness caused by methicillin-resistant Staphylococcus aureus. *Emerg Infect Dis,* 8**,** 82-4.

KAAS, R. S., LEEKITCHAROENPHON, P., AARESTRUP, F. M. & LUND, O. 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One,* 9**,** e104984.

KADLEC, K., EHRICHT, R., MONECKE, S., STEINACKER, U., KASPAR, H., MANKERTZ, J. & SCHWARZ, S. 2009. Diversity of antimicrobial resistance pheno- and genotypes of methicillin-resistant Staphylococcus aureus ST398 from diseased swine. *J Antimicrob Chemother,* 64**,** 1156-64.

KADLEC, K., FESSLE, A. T., HAUSCHILD, T. & SCHWARZ, S. 2012. Novel and uncommon antimicrobial resistance genes in livestock-associated methicillin-resistant Staphylococcus aureus. *Clinical Microbiology and Infection,* 18**,** 745-755.

KAHL, B. C. 2010. Impact of Staphylococcus aureus on the pathogenesis of chronic cystic fibrosis lung disease. *Int J Med Microbiol,* 300**,** 514-9.

KATAKWEBA, A. S., MUHAIRWA, A. P., ESPINOSA-GONGORA, C., GUARDABASSI, L., MTAMBO, M. M. & OLSEN, J. E. 2016. spa typing and antimicrobial resistance of Staphylococcus aureus from healthy humans, pigs and dogs in Tanzania. *J Infect Dev Ctries,* 10**,** 143-8.

KATAYAMA, Y., ITO, T. & HIRAMATSU, K. 2000. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother,* 44**,** 1549-55.

KIM, J., HONG, J., LIM, J.-A., HEU, S. & ROH, E. 2018. Improved multiplex PCR primers for rapid identification of coagulase-negative staphylococci. *Archives of Microbiology,* 200**,** 73-83.

KINABO, G. D., VAN DER VEN, A., MSUYA, L. J., SHAYO, A. M., SCHIMANA, W., NDARO, A., VAN ASTEN, H. A., DOLMANS, W. M., WARRIS, A. & HERMANS, P. W. 2013. Dynamics of nasopharyngeal bacterial colonisation in HIV-exposed young infants in Tanzania. *Trop Med Int Health,* 18**,** 286-95.

KLEIN, E., SMITH, D. L. & LAXMINARAYAN, R. 2007. Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. *Emerg Infect Dis,* 13**,** 1840-6.

KOS, V. N., DESJARDINS, C. A., GRIGGS, A., CERQUEIRA, G., VAN TONDER, A., HOLDEN, M. T., GODFREY, P., PALMER, K. L., BODI, K., MONGODIN, E. F., WORTMAN, J., FELDGARDEN, M., LAWLEY, T., GILL, S. R., HAAS, B. J., BIRREN, B. & GILMORE, M. S. 2012. Comparative genomics of vancomycin-resistant Staphylococcus aureus strains and their positions within the clade most commonly associated with Methicillin-resistant S. aureus hospital-acquired infection in the United States. *MBio,* 3.

KOYMANS, K. J., VRIELING, M., GORHAM, R. D. & VAN STRIJP, J. A. G. 2017. Staphylococcal Immune Evasion Proteins: Structure, Function, and Host Adaptation. *In:* BAGNOLI, F., RAPPUOLI, R. & GRANDI, G. (eds.) *Staphylococcus aureus: Microbiology, Pathology, Immunology, Therapy and Prophylaxis.* Cham: Springer International Publishing.

KURODA, M., OHTA, T., UCHIYAMA, I., BABA, T., YUZAWA, H., KOBAYASHI, I., CUI, L., OGUCHI, A., AOKI, K., NAGAI, Y., LIAN, J., ITO, T., KANAMORI, M., MATSUMARU, H., MARUYAMA, A., MURAKAMI, H., HOSOYAMA, A., MIZUTANI-UI, Y., TAKAHASHI, N. K., SAWANO, T., INOUE, R., KAITO, C., SEKIMIZU, K., HIRAKAWA, H., KUHARA, S., GOTO, S., YABUZAKI, J., KANEHISA, M., YAMASHITA, A., OSHIMA, K., FURUYA, K., YOSHINO, C., SHIBA, T., HATTORI, M., OGASAWARA, N., HAYASHI, H. & HIRAMATSU, K. 2001a. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet,* 357**,** 1225-40.

KURODA, M., OHTA, T., UCHIYAMA, I., BABA, T., YUZAWA, H., KOBAYASHI, I., CUI, L., OGUCHI, A., AOKI, K.-I., NAGAI, Y., LIAN, J., ITO, T., KANAMORI, M., MATSUMARU, H., MARUYAMA, A., MURAKAMI, H., HOSOYAMA, A., MIZUTANI-UI, Y., TAKAHASHI, N. K., SAWANO, T., INOUE, R.-I., KAITO, C., SEKIMIZU, K., HIRAKAWA, H., KUHARA, S., GOTO, S., YABUZAKI, J., KANEHISA, M., YAMASHITA, A., OSHIMA, K., FURUYA, K., YOSHINO, C., SHIBA, T., HATTORI, M., OGASAWARA, N., HAYASHI, H. & HIRAMATSU,

K. 2001b. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *The Lancet,* 357**,** 1225-1240.

KURTZ, S., PHILLIPPY, A., DELCHER, A. L., SMOOT, M., SHUMWAY, M., ANTONESCU, C. & SALZBERG, S. L. 2004. Versatile and open software for comparing large genomes. *Genome Biol,* 5**,** R12.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

LE LOIR, Y., BARON, F. & GAUTIER, M. 2003. Staphylococcus aureus and food poisoning. *Genet Mol Res,* 2**,** 63-76.

LEMMA, M. T., ZENEBE, Y., TULU, B., MEKONNEN, D. & MEKONNEN, Z. 2015. Methicillin Resistant Staphylococcus aureus among HIV Infected Pediatric Patients in Northwest Ethiopia: Carriage Rates and Antibiotic Co-Resistance Profiles. *PLoS One,* 10**,** e0137254.

LI, H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics,* 32**,** 2103-10.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LINA, G., BOHACH, G. A., NAIR, S. P., HIRAMATSU, K., JOUVIN-MARCHE, E. & MARIUZZA, R. 2004. Standard nomenclature for the superantigens expressed by Staphylococcus. *J Infect Dis,* 189**,** 2334-6.

LINA, G., PIEMONT, Y., GODAIL-GAMOT, F., BES, M., PETER, M. O., GAUDUCHON, V., VANDENESCH, F. & ETIENNE, J. 1999. Involvement of Panton-Valentine leukocidin-producing Staphylococcus aureus in primary skin infections and pneumonia. *Clin Infect Dis,* 29**,** 1128-32.

LINDERHOLM, A. 2016. Ancient DNA: the next generation – chapter and verse. *Biological Journal of the Linnean Society,* 117**,** 150-160.

LINDSAY, J. A. 2010. Genomic variation and evolution of Staphylococcus aureus. *International Journal of Medical Microbiology,* 300**,** 98-103.

LINDSAY, J. A. 2014. Staphylococcus aureus genomics and the impact of horizontal gene transfer. *International Journal of Medical Microbiology,* 304**,** 103-109.

LINDSAY, J. A. & HOLDEN, M. T. 2004. Staphylococcus aureus: superbug, super genome? *Trends Microbiol,* 12**,** 378-85.

LOEFFLER, A., BOAG, A. K., SUNG, J., LINDSAY, J. A., GUARDABASSI, L., DALSGAARD, A., SMITH, H., STEVENS, K. B. & LLOYD, D. H. 2005. Prevalence of methicillin-resistant Staphylococcus aureus among staff and pets in a small animal referral hospital in the UK. *J Antimicrob Chemother,* 56**,** 692-7.

LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEN, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol,* 30**,** 434-9.

LOWY, F. D. 1998. Staphylococcus aureus infections. *N Engl J Med,* 339**,** 520-32.

LOZANO, C., GHARSA, H., BEN SLAMA, K., ZARAZAGA, M. & TORRES, C. 2016a. Staphylococcus aureus in Animals and Food: Methicillin Resistance, Prevalence and Population Structure. A Review in the African Continent. *Microorganisms,* 4**,** 12.

LOZANO, C., GHARSA, H., BEN SLAMA, K., ZARAZAGA, M. & TORRES, C. 2016b. Staphylococcus aureus in Animals and Food: Methicillin Resistance, Prevalence and Population Structure. A Review in the African Continent. *Microorganisms,* 4.

MCCARTHY, A. J., LINDSAY, J. A. & LOEFFLER, A. 2012. Are all meticillin-resistant Staphylococcus aureus (MRSA) equal in all hosts? Epidemiological and genetic comparison between animal and human MRSA. *Veterinary Dermatology,* 23**,** 267-e54.

MEDIAVILLA, J. R., CHEN, L., UHLEMANN, A. C., HANSON, B. M., ROSENTHAL, M., STANAK, K., KOLL, B., FRIES, B. C., ARMELLINO, D., SCHILLING, M. E., WEISS, D., SMITH, T. C., LOWY, F. D. & KREISWIRTH, B. N. 2012. Methicillin-susceptible Staphylococcus aureus ST398, New York and New Jersey, USA. *Emerg Infect Dis,* 18**,** 700-2.

MEKA, V. G., PILLAI, S. K., SAKOULAS, G., WENNERSTEN, C., VENKATARAMAN, L., DEGIROLAMI, P. C., ELIOPOULOS, G. M., MOELLERING, R. C., JR. & GOLD, H. S. 2004. Linezolid resistance in sequential Staphylococcus aureus isolates associated with a T2500A mutation in the 23S rRNA gene and loss of a single copy of rRNA. *J Infect Dis,* 190**,** 311-7.

MELLMANN, A., BLETZ, S., BOKING, T., KIPP, F., BECKER, K., SCHULTES, A., PRIOR, K. & HARMSEN, D. 2016. Real-Time Genome Sequencing of Resistant Bacteria Provides Precision Infection Control in an Institutional Setting. *J Clin Microbiol,* 54**,** 2874-2881.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MONECKE, S., COOMBS, G., SHORE, A. C., COLEMAN, D. C., AKPAKA, P., BORG, M., CHOW, H., IP, M., JATZWAUK, L., JONAS, D., KADLEC, K., KEARNS, A., LAURENT, F., O'BRIEN, F. G., PEARSON, J., RUPPELT, A., SCHWARZ, S., SCICLUNA, E., SLICKERS, P., TAN, H.-L., WEBER, S. & EHRICHT, R. 2011. A Field Guide to Pandemic, Epidemic and Sporadic Clones of Methicillin-Resistant Staphylococcus aureus. *PLoS ONE,* 6**,** e17936.

MORADIGARAVAND, D., JAMROZY, D., MOSTOWY, R., ANDERSON, A., NICKERSON, E. K., THAIPADUNGPANIT, J., WUTHIEKANUN, V., LIMMATHUROTSAKUL, D., TANDHAVANANT, S., WIKRAIPHAT, C., WONGSUVAN, G., TEERAWATTANASOOK, N., JUTRAKUL, Y., SRISURAT, N., CHAIMANEE, P., EOIN WEST, T., BLANE, B., PARKHILL, J., CHANTRATITA, N. & PEACOCK, S. J. 2017. Evolution of the Staphylococcus argenteus ST2250 Clone in Northeastern Thailand Is Linked with the Acquisition of Livestock-Associated Staphylococcal Genes. *mBio,* 8**,** e00802-17.

MURCHAN, S., KAUFMANN, M. E., DEPLANO, A., DE RYCK, R., STRUELENS, M., ZINN, C. E., FUSSING, V., SALMENLINNA, S., VUOPIO-VARKILA, J., EL SOLH, N., CUNY, C., WITTE, W., TASSIOS, P. T., LEGAKIS, N., VAN LEEUWEN, W., VAN BELKUM, A., VINDEL, A., LACONCHA, I., GARAIZAR, J., HAEGGMAN, S., OLSSON-LILJEQUIST, B., RANSJO, U., COOMBES, G. & COOKSON, B. 2003. Harmonization of pulsed-field gel electrophoresis protocols

for epidemiological typing of strains of methicillin-resistant Staphylococcus aureus: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol,* 41**,** 1574-85.

NATHWANI, D., MORGAN, M., MASTERTON, R. G., DRYDEN, M., COOKSON, B. D., FRENCH, G. & LEWIS, D. 2008. Guidelines for UK practice for the diagnosis and management of methicillin-resistant Staphylococcus aureus (MRSA) infections presenting in the community. *J Antimicrob Chemother,* 61**,** 976-94.

NICKERSON, E. K., HONGSUWAN, M., LIMMATHUROTSAKUL, D., WUTHIEKANUN, V., SHAH, K. R., SRISOMANG, P., MAHAVANAKUL, W., WACHARAPRECHASGUL, T., FOWLER, V. G., WEST, T. E., TEERAWATANASUK, N., BECHER, H., WHITE, N. J., CHIERAKUL, W., DAY, N. P. & PEACOCK, S. J. 2009. Staphylococcus aureus bacteraemia in a tropical setting: patient outcome and impact of antibiotic resistance. *PLoS One,* 4**,** e4308.

NIELSEN, R. T., KEMP, M., HOLM, A., SKOV, M. N., DETLEFSEN, M., HASMAN, H., AARESTRUP, F. M., KAAS, R. S., NIELSEN, J. B., WESTH, H. & KOLMOS, H. J. 2016. Fatal Septicemia Linked to Transmission of MRSA Clonal Complex 398 in Hospital and Nursing Home, Denmark. *Emerging Infectious Diseases,* 22**,** 900-902.

NOBLE, W. C., VIRANI, Z. & CREE, R. G. A. 1992. Co-transfer of vancomycin and other resistance genes from Enterococcus faecalis NCTC 12201 to Staphylococcus aureus. *FEMS Microbiology Letters,* 93**,** 195-198.

NOVICK, R. P. & BOUANCHAUD, D. 1971. The problems of drug-resistant pathogenic bacteria. Extrachromosomal nature of drug resistance in Staphylococcus aureus. *Ann N Y Acad Sci,* 182**,** 279-94.

NOVICK, R. P., SCHLIEVERT, P. & RUZIN, A. 2001. Pathogenicity and resistance islands of staphylococci. *Microbes Infect,* 3**,** 585-94.

OGSTON, A. 1881. Report upon Micro-Organisms in Surgical Diseases. *British Medical Journal,* 1**,** 369-375.

OMUSE, G., VAN ZYL, K. N., HOEK, K., ABDULGADER, S., KARIUKI, S., WHITELAW, A. & REVATHI, G. 2016. Molecular characterization of Staphylococcus aureus isolates from various healthcare institutions in Nairobi,

Kenya: a cross sectional study. *Annals of Clinical Microbiology and Antimicrobials,* 15**,** 51.

OSTOJIC, M. & HUKIC, M. 2015. Genotypic and phenotypic characteristics of Methicillin-resistant Staphylococcus aureus (MRSA) strains, isolated on three different geography locations. *Bosn J Basic Med Sci,* 15**,** 48-56.

OUEDRAOGO, A.-S., DUNYACH-REMY, C., KISSOU, A., SANOU, S., PODA, A., KYELEM, C. G., SOLASSOL, J., BAÑULS, A.-L., VAN DE PERRE, P., OUÉDRAOGO, R., JEAN-PIERRE, H., LAVIGNE, J.-P. & GODREUIL, S. 2016. High Nasal Carriage Rate of Staphylococcus aureus Containing Panton-Valentine leukocidin- and EDIN-Encoding Genes in Community and Hospital Settings in Burkina Faso. *Frontiers in Microbiology,* 7**,** 1406.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T. G., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3693.

PAGE, A. J., TAYLOR, B., DELANEY, A. J., SOARES, J., SEEMANN, T., KEANE, J. A. & HARRIS, S. R. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom,* 2**,** e000056.

PANTUCEK, R., DOSKAR, J., RUZICKOVA, V., KASPAREK, P., ORACOVA, E., KVARDOVA, V. & ROSYPAL, S. 2004. Identification of bacteriophage types and their carriage in Staphylococcus aureus. *Arch Virol,* 149**,** 1689-703.

PEACOCK, S. J., DE SILVA, I. & LOWY, F. D. 2001. What determines nasal carriage of Staphylococcus aureus? *Trends in Microbiology,* 9**,** 605-610.

PETIT III, R. A. & READ, T. D. 2018. Staphylococcus aureus viewed from the perspective of 40,000+ genomes. *PeerJ Preprints,* 6**,** e26890v2.

PIGNATARI, A., PFALLER, M., HOLLIS, R., SESSO, R., LEME, I. & HERWALDT, L. 1990. Staphylococcus aureus colonization and infection in patients on continuous ambulatory peritoneal dialysis. *J Clin Microbiol,* 28**,** 1898-902.

PINCHUK, I. V., BESWICK, E. J. & REYES, V. E. 2010. Staphylococcal Enterotoxins. *Toxins,* 2**,** 2177-2197.

PLOUGH, H. H. 1945. Penicillin resistance of Staphylococcus aureus and its clinical implications. *Am J Clin Pathol,* 15**,** 446-51.

PRICE, J., CLAIRE GORDON, N., CROOK, D., LLEWELYN, M. & PAUL, J. 2013. The usefulness of whole genome sequencing in the management of Staphylococcus aureus infections. *Clinical Microbiology and Infection,* 19**,** 784-789.

PRICE, L. B., STEGGER, M., HASMAN, H., AZIZ, M., LARSEN, J., ANDERSEN, P. S., PEARSON, T., WATERS, A. E., FOSTER, J. T., SCHUPP, J., GILLECE, J., DRIEBE, E., LIU, C. M., SPRINGER, B., ZDOVC, I., BATTISTI, A., FRANCO, A., ZMUDZKI, J., SCHWARZ, S., BUTAYE, P., JOUY, E., POMBA, C., PORRERO, M. C., RUIMY, R., SMITH, T. C., ROBINSON, D. A., WEESE, J. S., ARRIOLA, C. S., YU, F., LAURENT, F., KEIM, P., SKOV, R. & AARESTRUP, F. M. 2012a. Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio,* 3.

PRICE, L. B., STEGGER, M., HASMAN, H., AZIZ, M., LARSEN, J., ANDERSEN, P. S., PEARSON, T., WATERS, A. E., FOSTER, J. T., SCHUPP, J., GILLECE, J., DRIEBE, E., LIU, C. M., SPRINGER, B., ZDOVC, I., BATTISTI, A., FRANCO, A., ŻMUDZKI, J., SCHWARZ, S., BUTAYE, P., JOUY, E., POMBA, C., PORRERO, M. C., RUIMY, R., SMITH, T. C., ROBINSON, D. A., WEESE, J. S., ARRIOLA, C. S., YU, F., LAURENT, F., KEIM, P., SKOV, R. & AARESTRUP, F. M. 2012b. Staphylococcus aureus CC398: Host Adaptation and Emergence of Methicillin Resistance in Livestock. *mBio,* 3**,** e00305-11.

PRUITT, K. D., TATUSOVA, T., BROWN, G. R. & MAGLOTT, D. R. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res,* 40**,** D130-5.

QUAIL, M. A., KOZAREWA, I., SMITH, F., SCALLY, A., STEPHENS, P. J., DURBIN, R., SWERDLOW, H. & TURNER, D. J. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods,* 5**,** 1005-10.

RAJENDRAN, P. M., YOUNG, D., MAURER, T., CHAMBERS, H., PERDREAU-REMINGTON, F., RO, P. & HARRIS, H. 2007. Randomized, double-blind, placebo-controlled trial of cephalexin for treatment of uncomplicated skin abscesses in a population at risk for community-acquired methicillin-resistant Staphylococcus aureus infection. *Antimicrob Agents Chemother,* 51**,** 4044-8.

RAMMELKAMP, C. H. & MAXON, T. 1942. Resistance of Staphylococcus aureus to the Action of Penicillin. *Proceedings of the Society for Experimental Biology and Medicine,* 51**,** 386-389.

REGENTHAL, P., HANSEN, J. S., ANDRÉ, I. & LINDKVIST-PETERSSON, K. 2017. Thermal stability and structural changes in bacterial toxins responsible for food poisoning. *PLoS ONE,* 12**,** e0172445.

REUTER, S., TOROK, M. E., HOLDEN, M. T., REYNOLDS, R., RAVEN, K. E., BLANE, B., DONKER, T., BENTLEY, S. D., AANENSEN, D. M., GRUNDMANN, H., FEIL, E. J., SPRATT, B. G., PARKHILL, J. & PEACOCK, S. J. 2016. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Res,* 26**,** 263-70.

RICHARDSON, E. J., BACIGALUPE, R., HARRISON, E. M., WEINERT, L. A., LYCETT, S., VRIELING, M., ROBB, K., HOSKISSON, P. A., HOLDEN, M. T. G., FEIL, E. J., PATERSON, G. K., TONG, S. Y. C., SHITTU, A., VAN WAMEL, W., AANENSEN, D. M., PARKHILL, J., PEACOCK, S. J., CORANDER, J., HOLMES, M. & FITZGERALD, J. R. 2018. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat Ecol Evol*.

RINSKY, J. L., NADIMPALLI, M., WING, S., HALL, D., BARON, D., PRICE, L. B., LARSEN, J., STEGGER, M., STEWART, J. & HEANEY, C. D. 2013. Livestock-associated methicillin and multidrug resistant Staphylococcus aureus is present among industrial, not antibiotic-free livestock operation workers in North Carolina. *PLoS One,* 8**,** e67641.

RUHE, J. J. & MENON, A. 2007. Tetracyclines as an oral treatment option for patients with community onset skin and soft tissue infections caused by methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother,* 51**,** 3298-303.

SALLAM, K. I., ABD-ELGHANY, S. M., ELHADIDY, M. & TAMURA, T. 2015. Molecular Characterization and Antimicrobial Resistance Profile of Methicillin-Resistant Staphylococcus aureus in Retail Chicken. *J Food Prot,* 78**,** 1879-84.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1992. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology,* 24**,** 104-8.

SCHADT, E. E., TURNER, S. & KASARSKIS, A. 2010. A window into third-generation sequencing. *Hum Mol Genet,* 19**,** R227-40.

SCHAUMBURG, F., ALABI, A. S., KOCK, R., MELLMANN, A., KREMSNER, P. G., BOESCH, C., BECKER, K., LEENDERTZ, F. H. & PETERS, G. 2012. Highly divergent Staphylococcus aureus isolates from African non-human primates. *Environ Microbiol Rep,* 4**,** 141-6.

SCHAUMBURG, F., ALABI, A. S., PETERS, G. & BECKER, K. 2014a. New epidemiology of Staphylococcus aureus infection in Africa. *Clin Microbiol Infect,* 20**,** 589-96.

SCHAUMBURG, F., ALABI, A. S., PETERS, G. & BECKER, K. 2014b. New epidemiology of Staphylococcus aureus infection in Africa. *Clinical Microbiology and Infection,* 20**,** 589-596.

SCHAUMBURG, F., MUGISHA, L., KAPPELLER, P., FICHTEL, C., KÖCK, R., KÖNDGEN, S., BECKER, K., BOESCH, C., PETERS, G. & LEENDERTZ, F. 2013. Evaluation of Non-Invasive Biological Samples to Monitor Staphylococcus aureus Colonization in Great Apes and Lemurs. *PLoS ONE,* 8**,** e78046.

SCHAUMBURG, F., PAULY, M., ANOH, E., MOSSOUN, A., WIERSMA, L., SCHUBERT, G., FLAMMEN, A., ALABI, A. S., MUYEMBE-TAMFUM, J.-J., GROBUSCH, M. P., KARHEMERE, S., AKOUA-KOFFI, C., COUACY-HYMANN, E., KREMSNER, P. G., MELLMANN, A., BECKER, K., LEENDERTZ, F. H. & PETERS, G. 2015. Staphylococcus aureus complex from animals and humans in three remote African regions. *Clinical Microbiology and Infection,* 21**,** 345.e1-345.e8.

SCHIJFFELEN, M. J., BOEL, C. H., VAN STRIJP, J. A. & FLUIT, A. C. 2010. Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis. *BMC Genomics,* 11**,** 376.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SENGHORE, M., BAYLISS, S. C., KWAMBANA-ADAMS, B. A., FOSTER-NYARKO, E., MANNEH, J., DIONE, M., BADJI, H., EBRUKE, C., DOUGHTY, E. L., THORPE, H. A., JASINSKA, A. J., SCHMITT, C. A., CRAMER, J. D., TURNER,

T. R., WEINSTOCK, G., FREIMER, N. B., PALLEN, M. J., FEIL, E. J. & ANTONIO, M. 2016. Transmission of Staphylococcus aureus from Humans to Green Monkeys in The Gambia as Revealed by Whole-Genome Sequencing. *Appl Environ Microbiol,* 82**,** 5910-7.

SHEPHEARD, M. A., FLEMING, V. M., CONNOR, T. R., CORANDER, J., FEIL, E. J., FRASER, C. & HANAGE, W. P. 2013. Historical zoonoses and other changes in host tropism of Staphylococcus aureus, identified by phylogenetic analysis of a population dataset. *PLoS One,* 8**,** e62369.

SIDDIQUI, A. H. & WHITTEN, R. A. 2018. Methicillin Resistant Staphylococcus Aureus (MRSA). *StatPearls.* Treasure Island (FL): StatPearls Publishing

StatPearls Publishing LLC.

SMITH, D. S., SIGGINS, M. K., GIERULA, M., PICHON, B., TURNER, C. E., LYNSKEY, N. N., MOSAVIE, M., KEARNS, A. M., EDWARDS, R. J. & SRISKANDAN, S. 2016. Identification of commonly expressed exoproteins and proteolytic cleavage events by proteomic mining of clinically relevant UK isolates of Staphylococcus aureus. *Microb Genom,* 2**,** e000049.

SOGE, O. O., NO, D., MICHAEL, K. E., DANKOFF, J., LANE, J., VOGEL, K., SMEDLEY, J. & ROBERTS, M. C. 2016. Transmission of MDR MRSA between primates, their environment and personnel at a United States primate centre. *Journal of Antimicrobial Chemotherapy,* 71**,** 2798-2803.

SONG, M., SHI, C., XU, X. & SHI, X. 2016. Molecular Typing and Virulence Gene Profiles of Enterotoxin Gene Cluster (egc)-Positive Staphylococcus aureus Isolates Obtained from Various Food and Clinical Specimens. *Foodborne Pathog Dis,* 13**,** 592-601.

STEGGER, M., ANDERSEN, P. S., KEARNS, A., PICHON, B., HOLMES, M. A., EDWARDS, G., LAURENT, F., TEALE, C., SKOV, R. & LARSEN, A. R. 2012. Rapid detection, differentiation and typing of methicillin-resistant Staphylococcus aureus harbouring either mecA or the new mecA homologue mecA(LGA251). *Clin Microbiol Infect,* 18**,** 395-400.

STRAUS, L., RUFFING, U., ABDULLA, S., ALABI, A., AKULENKO, R., GARRINE, M., GERMANN, A., GROBUSCH, M. P., HELMS, V., HERRMANN, M., KAZIMOTO, T., KERN, W., MANDOMANDO, I., PETERS, G.,

SCHAUMBURG, F., VON MÜLLER, L. & MELLMANN, A. 2016. Detecting Staphylococcus aureus Virulence and Resistance Genes: a Comparison of Whole-Genome Sequencing and DNA Microarray Technology. *Journal of Clinical Microbiology,* 54**,** 1008-1016.

SUNG, J. M.-L., LLOYD, D. H. & LINDSAY, J. A. 2008. Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology,* 154**,** 1949-1959.

SUTHERLAND, R., CROYDON, E. A. P. & ROLINSON, G. N. 1970. Flucloxacillin, a New Isoxazolyl Penicillin, Compared with Oxacillin, Cloxacillin, and Dicloxacillin. *British Medical Journal,* 4**,** 455-460.

SVENSSON, K., HELLMARK, B. & SÖDERQUIST, B. O. 2011. Characterization of SCCmec elements in methicillin-resistant Staphylococcus epidermidis isolated from blood cultures from neonates during three decades. *APMIS,* 119**,** 885-893.

TACCONELLI, E., CARRARA, E., SAVOLDI, A., HARBARTH, S., MENDELSON, M., MONNET, D. L., PULCINI, C., KAHLMETER, G., KLUYTMANS, J., CARMELI, Y., OUELLETTE, M., OUTTERSON, K., PATEL, J., CAVALERI, M., COX, E. M., HOUCHENS, C. R., GRAYSON, M. L., HANSEN, P., SINGH, N., THEURETZBACHER, U. & MAGRINI, N. 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis,* 18**,** 318-327.

TAYLOR, L. H., LATHAM, S. M. & WOOLHOUSE, M. E. 2001. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 983-9.

THAKKER, M., PARK, J. S., CAREY, V. & LEE, J. C. 1998. Staphylococcus aureus serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. *Infect Immun,* 66**,** 5183-9.

TOSAS AUGUET, O., BETLEY, J. R., STABLER, R. A., PATEL, A., IOANNOU, A., MARBACH, H., HEARN, P., ARYEE, A., GOLDENBERG, S. D., OTTER, J. A., DESAI, N., KARADAG, T., GRUNDY, C., GAUNT, M. W., COOPER, B. S., EDGEWORTH, J. D. & KYPRAIOS, T. 2016. Evidence for Community Transmission of Community-Associated but Not Health-Care-Associated Methicillin-Resistant Staphylococcus Aureus Strains Linked to Social and Material

Deprivation: Spatial Analysis of Cross-sectional Data. *PLoS Medicine,* 13**,** e1001944.

TSIODRAS, S., GOLD, H. S., SAKOULAS, G., ELIOPOULOS, G. M., WENNERSTEN, C., VENKATARAMAN, L., MOELLERING, R. C. & FERRARO, M. J. 2001. Linezolid resistance in a clinical isolate of Staphylococcus aureus. *The Lancet,* 358**,** 207-208.

UHLEMANN, A. C., HAFER, C., MIKO, B. A., SOWASH, M. G., SULLIVAN, S. B., SHU, Q. & LOWY, F. D. 2013. Emergence of sequence type 398 as a community- and healthcare-associated methicillin-susceptible staphylococcus aureus in northern Manhattan. *Clin Infect Dis,* 57**,** 700-3.

URWIN, R. & MAIDEN, M. C. 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol,* 11**,** 479-87.

VAN BELKUM, A., MELLES, D. C., NOUWEN, J., VAN LEEUWEN, W. B., VAN WAMEL, W., VOS, M. C., WERTHEIM, H. F. & VERBRUGH, H. A. 2009. Co-evolutionary aspects of human colonisation and infection by Staphylococcus aureus. *Infect Genet Evol,* 9**,** 32-47.

VAN BELKUM, A., MELLES, D. C., SNIJDERS, S. V., VAN LEEUWEN, W. B., WERTHEIM, H. F. L., NOUWEN, J. L., VERBRUGH, H. A. & ETIENNE, J. 2006. Clonal Distribution and Differential Occurrence of the Enterotoxin Gene Cluster, egc, in Carriage- versus Bacteremia-Associated Isolates of Staphylococcus aureus. *Journal of Clinical Microbiology,* 44**,** 1555-1557.

VAN CLEEF, B. A., GRAVELAND, H., HAENEN, A. P., VAN DE GIESSEN, A. W., HEEDERIK, D., WAGENAAR, J. A. & KLUYTMANS, J. A. 2011. Persistence of livestock-associated methicillin-resistant Staphylococcus aureus in field workers after short-term occupational exposure to pigs and veal calves. *J Clin Microbiol,* 49**,** 1030-3.

VAN DUIJKEREN, E., JANSEN, M. D., FLEMMING, S. C., DE NEELING, H., WAGENAAR, J. A., SCHOORMANS, A. H., VAN NES, A. & FLUIT, A. C. 2007. Methicillin-resistant Staphylococcus aureus in pigs with exudative epidermitis. *Emerg Infect Dis,* 13**,** 1408-10.

VAN WAMEL, W. J., ROOIJAKKERS, S. H., RUYKEN, M., VAN KESSEL, K. P. & VAN STRIJP, J. A. 2006a. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages. *J Bacteriol,* 188**,** 1310-5.

VAN WAMEL, W. J. B., ROOIJAKKERS, S. H. M., RUYKEN, M., VAN KESSEL, K. P. M. & VAN STRIJP, J. A. G. 2006b. The Innate Immune Modulators Staphylococcal Complement Inhibitor and Chemotaxis Inhibitory Protein of Staphylococcus aureus Are Located on β-Hemolysin-Converting Bacteriophages. *Journal of Bacteriology,* 188**,** 1310-1315.

VERKAIK, N. J., BENARD, M., BOELENS, H. A., DE VOGEL, C. P., NOUWEN, J. L., VERBRUGH, H. A., MELLES, D. C., VAN BELKUM, A. & VAN WAMEL, W. J. 2011a. Immune evasion cluster-positive bacteriophages are highly prevalent among human Staphylococcus aureus strains, but they are not essential in the first stages of nasal colonization. *Clin Microbiol Infect,* 17**,** 343-8.

VERKAIK, N. J., BENARD, M., BOELENS, H. A., DE VOGEL, C. P., NOUWEN, J. L., VERBRUGH, H. A., MELLES, D. C., VAN BELKUM, A. & VAN WAMEL, W. J. B. 2011b. Immune evasion cluster-positive bacteriophages are highly prevalent among human Staphylococcus aureus strains, but they are not essential in the first stages of nasal colonization. *Clinical Microbiology and Infection,* 17**,** 343-348.

VOS, M. C., OTT, A. & VERBRUGH, H. A. 2005. Successful Search-and-Destroy Policy for Methicillin-Resistant Staphylococcus aureus in The Netherlands. *Journal of Clinical Microbiology,* 43**,** 2034.

VOSS, A., LOEFFEN, F., BAKKER, J., KLAASSEN, C. & WULF, M. 2005. Methicillin-resistant Staphylococcus aureus in pig farming. *Emerg Infect Dis,* 11**,** 1965-6.

WALTHER, B., WIELER, L. H., FRIEDRICH, A. W., HANSSEN, A.-M., KOHN, B., BRUNNBERG, L. & LÜBKE-BECKER, A. 2008. Methicillin-resistant Staphylococcus aureus (MRSA) isolated from small and exotic animals at a university hospital during routine microbiological examinations. *Veterinary Microbiology,* 127**,** 171-178.

WANG, X., TOWERS, S., PANCHANATHAN, S. & CHOWELL, G. 2013. A population based study of seasonality of skin and soft tissue infections: implications for the spread of CA-MRSA. *PLoS One,* 8**,** e60872.

WANG, Y., LIU, Q., LIU, Q., GAO, Q., LU, H., MENG, H., XIE, Y., HUANG, Q., MA, X., WANG, H., QIN, J., LI, Q., LI, T., XIA, Q. & LI, M. 2018. Phylogenetic analysis and virulence determinant of the host-adapted Staphylococcus aureus lineage ST188 in China. *Emerging Microbes & Infections,* 7**,** 45.

WARNE, B., HARKINS, C. P., HARRIS, S. R., VATSIOU, A., STANLEY-WALL, N., PARKHILL, J., PEACOCK, S. J., PALMER, T. & HOLDEN, M. T. 2016. The Ess/Type VII secretion system of Staphylococcus aureus shows unexpected genetic diversity. *BMC Genomics,* 17**,** 222.

WEINKE, T., SCHILLER, R., FEHRENBACH, F. J. & POHLE, H. D. 1992. Association between Staphylococcus aureus nasopharyngeal colonization and septicemia in patients infected with the human immunodeficiency virus. *Eur J Clin Microbiol Infect Dis,* 11**,** 985-9.

WEINSTEIN, H. J. 1959. The Relation between the Nasal-Staphylococcal-Carrier State and the Incidence of Postoperative Complications. *New England Journal of Medicine,* 260**,** 1303-1308.

WERTHEIM, H. F., VOS, M. C., BOELENS, H. A., VOSS, A., VANDENBROUCKE-GRAULS, C. M., MEESTER, M. H., KLUYTMANS, J. A., VAN KEULEN, P. H. & VERBRUGH, H. A. 2004. Low prevalence of methicillin-resistant Staphylococcus aureus (MRSA) at hospital admission in the Netherlands: the value of search and destroy and restrictive antibiotic use. *J Hosp Infect,* 56**,** 321-5.

WOOLHOUSE, M., WARD, M., VAN BUNNIK, B. & FARRAR, J. 2015. Antimicrobial resistance in humans, livestock and the wider environment. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 370**,** 20140083.

WOOLHOUSE, M. E., HAYDON, D. T. & ANTIA, R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol Evol,* 20**,** 238-44.

XIA, G. & WOLZ, C. 2014. Phages of Staphylococcus aureus and their impact on host evolution. *Infection, Genetics and Evolution,* 21**,** 593-601.

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**,** 6637-42.

YARWOOD, J. M. & SCHLIEVERT, P. M. 2003. Quorum sensing in Staphylococcus infections. *The Journal of Clinical Investigation,* 112**,** 1620-1625.

ZANKARI, E., HASMAN, H., COSENTINO, S., VESTERGAARD, M., RASMUSSEN, S., LUND, O., AARESTRUP, F. M. & LARSEN, M. V. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother,* 67**,** 2640-4.

ZECCONI, A. & SCALI, F. 2013. Staphylococcus aureus virulence factors in evasion from innate immune defenses in human and animal diseases. *Immunology Letters,* 150**,** 12-22.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.

ZHANG, S., SUN, X., CHANG, W., DAI, Y. & MA, X. 2015. Systematic Review and Meta-Analysis of the Epidemiology of Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate Staphylococcus aureus Isolates. *PLoS One,* 10**,** e0136082.

ZHAO, H., HU, F., JIN, S., XU, X., ZOU, Y., DING, B., HE, C., GONG, F. & LIU, Q. 2016. Typing of Panton-Valentine Leukocidin-Encoding Phages and lukSF-PV Gene Sequence Variation in Staphylococcus aureus from China. *Frontiers in Microbiology,* 7**,** 1200.

APPENDICES:

Supplementary file 1: Table showing MLST results with loci number of seven housekeeping genes based on assemblies

ISOLATE	ST	NEW ST	CONTAMINATION	ARC	AROE	GLPF	GMK	PTA	TPI	YQIL
24977_8#292	6			12	4	1	4	12	1	3
24977_8#293	580			3	35	48	19	20	26	39
24977_8#294	25			4	1	4	1	5	5	4
24977_8#310	6			12	4	1	4	12	1	3
24977_8#314	25			4	1	4	1	5	5	4
24977_8#313	25			4	1	4	1	5	5	4
24977_8#323	789			3	4	1	4	4	6	3
24977_8#326	188			3	1	1	8	1	1	1
24977_8#327	580			3	35	48	19	20	26	39
24977_8#328	15			13	13	1	1	12	11	13
24977_8#330	188			3	1	1	8	1	1	1
24977_8#333	152			46	75	49	44	13	68	60
24977_8#335	97			3	1	1	1	1	5	3
24977_8#336	152			46	75	49	44	13	68	60
24977_8#342	188			3	1	1	8	1	1	1
24977_8#346	22			7	6	1	5	8	8	6
24977_8#353	789			3	4	1	4	4	6	3
24977_8#356	188			3	1	1	8	1	1	1
24977_8#368	188			3	1	1	8	1	1	1
24977_8#371	188			3	1	1	8	1	1	1
24977_8#374	789			3	4	1	4	4	6	3
24977_8#380	580			3	35	48	19	20	26	39

24977_8#383	22			7	6	1	5	8	8	6

Supplementary file 2: Metadata of the sample isolates

ID	Source (Human/Pig)	Homestead Number	Rearing Type	Size Of Farm (Acres)	Herd Size	Age Of Handler	Chickens	Cattle	Recent Mastitis	Sheep	Goats	Pets (Cats)
24977_8#292	Human	22	Zero Grazing	5	21	21	Yes	No	No	No	No	No
24977_8#293	Swine	21	Zero Grazing	2.1	2	21	Yes	Yes	No	No	No	Yes
24977_8#294	Swine	3	Zero Grazing	2	43	21	Yes	No	No	Yes	Yes	No
24977_8#310	Human	5	Zero Grazing	4	25	43	Yes	No	No	No	No	No
24977_8#313	Human	6	Zero Grazing	2	32	12	Yes	No	No	No	Yes	No
24977_8#314	Swine	8	Zero Grazing	1	6	33	Yes	No	No	No	No	Yes
24977_8#323	Swine	4	Zero Grazing	3	32	32	Yes	No	No	No	No	Yes
24977_8#326	Human	33	Zero Grazing	1	8	31	Yes	No	No	No	No	Yes
24977_8#327	Swine	13	Free Range	2	12	23	No	No	No	No	Yes	No
24977_8#328	Human	24	Zero Grazing	0.6	3	32	Yes	Yes	Yes	No	No	No
24977_8#330	Swine	24	Zero Grazing	0.6	9	32	Yes	Yes	Yes	No	No	No
24977_8#333	Human	90	Zero Grazing	1	4	32	No	No	No	No	No	No
24977_8#335	Human	5	Zero Grazing	4	25	43	Yes	No	No	No	No	No
24977_8#336	Swine	67	Zero Grazing	2	7	43	Yes	No	No	No	Yes	Yes
24977_8#342	Swine	45	Zero Grazing	1.5	6	32	Yes	Yes	Yes	No	No	No
24977_8#346	Swine	6	Zero Grazing	2	32	12	Yes	No	No	No	Yes	No
24977_8#353	Human	2	Zero Grazing	0.5	6	32	Yes	No	No	No	No	Yes
24977_8#356	Human	33	Zero Grazing	1	8	31	Yes	No	No	No	No	No
24977_8#368	Swine	21	Zero Grazing	2.1	2	21	Yes	Yes	No	No	No	Yes
24977_8#371	Swine	19	Intensive	2	125	21	Yes	Yes	Yes	Yes	Yes	No

24977_8#374	HUMAN	104	ZERO GRAZING	1	7	24	YES	NO	NO	NO	NO	YES
24977_8#380	HUMAN	13	FREE RANGE	2	12	23	YES	NO	NO	NO	YES	NO
24977_8#383	HUMAN	1	FREE RANGE	1	6	32	YES	YES	NO	NO	YES	YES

Supplementary file 3: In vitro antibiotic susceptibility test of 23 strains against 17 panels of antibiotic drugs

Id	OFX	INTP	F	INTP	LZD	INTP	QD	INTP	AML	INTP	IPM	INTP	E	INTP	DO	INTP
24977_8#292	26	S	6	R	6	R	6	R			30		6	R	6	R
24977_8#293	28	S	24	S	24	S	6	R	22		24		28	S	20	S
24977_8#294	24	S	23	S	26	S	28	S	20		20		6	R	28	S
24977_8#310	6	R	20	S	18	R	14	R	8		20		6	R	12	R
24977_8#313	30	S	28	S	24	S	20	S	20		24		6	R	22	S
24977_8#314			20	S	28	S	22	S	18		24		20	I	8	R
24977_8#323	26	S	20	S	22	S	16	I	10		26		20	I	22	S
24977_8#326			24	S	22	S	20	S	20		30		20	I	28	S
24977_8#327	20	S	6	R	22	S	6	R	8		36		20	S	8	R
24977_8#328			20	S	12	R	24	S	16		28		20	I	24	S
24977_8#330	30	S	26	S	36	S	24	S	19		30		18	I	10	R
24977_8#333	22	S	20	S	22	S	18	I	28		30		6	R	30	S
24977_8#335	20	S	22	S	30	S	28	S	20		26		6	R	20	S
24977_8#336	6	R	24	S	32	S	30	S	18		24		22	I		
24977_8#342	28	S	24	S	36	S	28	S	18		24		6	R	20	S
24977_8#346	26	S	22	S	24	S	6	R	6		32		30	S	18	S
24977_8#353			22	S	24	S	22	S	18		28		8	R		
24977_8#356	26	S	30	S	22	S	24	S	30		30		28	S		
24977_8#368			26	S	32	S	20	S	24		24		24	S		
24977_8#371	24	S	24	S	34	S	24	S	30		34		18	I		
24977_8#374	22	S	22	S	10	R	30	S	34		28		20	I		
24977_8#380			20	S	36	S	24	S	18		30		22	I		
24977_8#383	20	S	28	S	30	S	16	I	18		26		18	I		

ID	AMC	INTP	AMP	INTP	CAZ	INTP	SXT	INTP	CN	INTP	C	INTP	CIP	INTP	NA	INTP	FOX	INTP
24977_8#292	30	S	18	R	18	S	22	S	14	I	24	S	29	S	6		26	S
24977_8#293	30	S	20	R	20	S	22	S	22	S	24	S	28	S	10		28	S
24977_8#294	25	S	15	R	15	I	6	R	20	S	24	S	34	S	6		28	S
24977_8#310	20	S	10	R	6	R	30	S	8	R	8	R	18	I	6		6	R
24977_8#313	22	S	12	R	20	S	24	S	18	S	22	S	20	I	8		26	S
24977_8#314	20	S	8	R	6	R	6	R	6	R	16	I	8	R	6		24	S
24977_8#323	24	S	14	R	20	S	22	S	22	S	20	S	22	S	6		18	R
24977_8#326	28	S	10	R	20	S	26	S	22	S	24	S	26	S	8		28	S
24977_8#327	22	S	18	R	20	S	6	R	18	S	20	S	20	I	30		6	R
24977_8#328	20	S	12	R	18	S	10	R	24	S	26	S	26	S	8		26	S
24977_8#330	22	S	20	R	28	S	30	S	20	S	22	S	24	S	10		34	S
24977_8#333	20	S	6	R	6	R	6	R	14	I	20	S	20	I	14		24	S
24977_8#335	24	S	18	R	18	S	10	R	22	S	8	R	26	S	14		34	S
24977_8#336	26	S	14	R	18	S	32	S	24	S	15	I	22	S	12		30	S
24977_8#342	26	S	20	R	20	S	12	I	20	S	26	S	30	S	6		36	S
24977_8#346	34	S	32	S	18	S	28	S	26	S	28	S	29	S	10		28	S
24977_8#353	24	S	30	S	22	S	22	S	12	R	6	R	28	S	14		6	R
24977_8#356	26	S	22	R	18	S	24	S	24	S	24	S	24	S	16		24	S
24977_8#368	22	S	24	R	20	S	26	S	20	S	30	S	30	S	22		22	S
24977_8#371	20	S	22	R	26	S	26	S	18	S	24	S	34	S	14		30	S
24977_8#374	30	S	26	R	24	S	22	S	10	R	26	S	32	S	16		32	S
24977_8#380	24	S	20	R	22	S	20	S	16	S	28	S	22	S	18		24	S
24977_8#383	22	S	28	R	10	S	18	S	14	I	16	I	28	S	8		26	S

Supplementary file 5: Comparison of phenotypic testing of 23 strains with genotypic resistance genes based on Resfinder and Argannot results. False positive indicates presence of AMR gene in WGS but absence in phenotypic results while False negative indicates absence of AMR gene in WGS but positive in phenotypic results

Antibiotic Drug	Antibiotic Class	Prevalence (Phenotypic)	Resistance Genes	False Negative	False Positive	Discordance Total	Concordance Proportions
Ampicillin	β-Lactams	21	Blaz	1	2	3	40/43
Gentamycin	Aminoglycosides	4	Aph(2), Aadc	2	1	3	4/7
Trimethoprim	Folate Synthesis Inhibitors	6	Dfrg	3	1	4	6/10
Ciprofloxacin	Quinolones	1	Grla (S80f), Gyra (S84l)	1	0	1	0/1
Erythromycin	Macrolides	8	Ermc, Erm33, Ermb	5	1	6	10/16
Cefoxitin	Low Affinity Pbp2	3	Meca	2	0	3	1 / 4
Total		N=39		Error N= 14/39 35.89%	Error N= 5/42 11.9%	N= 20/81 24.7 %	N= 61/81 75.30%

.

Supplementary 7: Number of shared accessory genes between ST580 and ST398

	MS SA	MS SA	MS SA	MS SA	MS SA	MS SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	SA	MS SA	MS SA	MS SA	MS SA	MS SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA	MR SA
T1176	299	297	296	69	94	115	68	63	57	58	69	69	63	61	63	56	49	27	37	35	102	100	115	99	78	38	52	40	29	65	56	62	55
T1176	297	337	296	69	94	115	68	63	57	58	69	69	63	61	63	56	49	27	37	35	102	100	115	99	78	38	52	40	29	65	56	62	55
T1176	296	296	298	69	94	115	68	63	57	58	69	69	63	61	62	56	49	27	37	34	102	100	115	99	77	37	52	40	28	65	56	62	55
T034	69	69	69	371	176	120	198	198	181	173	195	197	193	179	216	183	165	123	153	13	122	146	121	151	65	147	171	149	119	157	190	246	172
T571	94	94	94	176	434	146	165	159	147	154	176	174	162	157	175	149	135	106	119	11	133	137	145	134	95	126	157	136	115	150	166	166	132
T1451	115	115	115	120	146	222	121	117	101	107	129	127	117	105	121	104	93	62	69	71	187	185	203	185	141	77	107	89	59	118	110	112	88
T011	68	68	68	198	165	121	305	295	271	201	252	253	280	261	225	233	218	185	183	20	123	123	118	123	70	173	219	191	148	185	248	261	173
T011	63	63	63	198	159	117	295	305	269	201	247	246	275	256	223	229	218	183	181	20	121	119	115	122	70	172	215	187	148	181	244	259	169
T011	57	57	57	181	147	101	271	269	299	196	226	225	259	249	209	221	210	186	176	20	111	110	103	113	63	170	209	187	152	165	239	244	165
T034	58	58	58	173	154	107	201	201	196	418	233	235	195	193	211	179	165	147	156	16	106	104	109	106	73	155	199	188	124	151	189	183	154
T034	69	69	69	195	176	129	252	247	226	233	301	286	242	226	255	210	188	161	169	17	124	125	124	124	70	175	231	210	143	195	221	223	171
T034	69	69	69	197	174	127	253	246	225	235	286	293	242	227	251	212	189	162	172	17	123	123	126	123	73	176	233	207	138	195	223	225	169
T2741	63	63	63	193	162	117	280	275	259	195	242	242	327	296	222	229	211	177	188	19	117	117	115	117	65	185	225	197	149	175	241	265	176
T2741	61	61	61	179	157	105	261	256	249	193	226	227	296	322	214	223	206	181	191	19	109	108	104	108	65	182	216	197	151	165	236	252	171
T034	63	63	62	216	175	121	225	223	209	211	255	251	222	214	366	197	181	150	168	16	120	144	119	150	71	179	202	187	171	203	213	236	156
T2383	56	56	56	183	149	104	233	229	221	179	210	212	229	223	197	305	248	163	194	17	111	114	108	113	67	166	209	180	150	160	218	239	158
T011	49	49	49	165	135	93	218	218	210	165	188	189	211	206	181	248	303	173	191	18	95	95	96	95	66	162	187	168	151	145	208	223	150
T011	27	27	27	123	106	62	185	183	186	147	161	162	177	181	150	163	173	235	166	18	64	64	62	66	59	136	167	162	147	112	187	184	123
T011	37	37	37	153	119	69	183	181	176	156	169	172	188	191	168	194	191	166	296	16	69	74	68	71	63	159	168	167	148	121	192	197	146
T011	35	35	34	137	115	71	200	200	203	162	173	175	198	196	162	179	182	188	167	24	75	75	74	77	57	151	173	165	145	124	197	199	135
T571	102	102	102	122	133	187	123	121	111	106	124	123	117	109	120	111	95	64	69	75	217	206	193	199	139	80	111	93	64	124	118	119	95
T571	100	100	100	146	137	185	123	119	110	104	125	123	117	108	144	114	95	64	74	75	206	334	189	207	140	89	111	93	100	150	120	139	98
T571	115	115	115	121	145	203	118	115	103	109	124	126	115	104	119	108	96	62	68	74	193	189	227	191	145	77	108	87	61	117	110	112	88
T571	99	99	99	151	134	185	123	122	113	106	124	123	117	108	150	113	95	66	71	77	199	207	191	277	130	79	107	88	70	129	118	142	95
T2383	78	78	77	65	95	141	70	70	63	73	70	73	65	65	71	67	66	59	63	57	139	140	145	130	202	67	70	64	64	69	70	72	58
T034	38	38	37	147	126	77	173	172	170	155	175	176	185	182	179	166	162	136	159	15	80	89	77	79	67	292	174	164	148	142	179	194	140
T034	52	52	52	171	157	107	219	215	209	199	231	233	225	216	202	209	187	167	168	17	111	111	108	107	70	174	277	220	137	168	221	223	156
T034	40	40	40	149	136	89	191	187	187	188	210	207	197	197	187	180	168	162	167	16	93	93	87	88	64	164	220	258	139	147	198	192	142
T011	29	29	28	119	115	59	148	148	152	124	143	138	149	151	171	150	151	147	148	14	64	100	61	70	64	148	137	139	436	157	188	184	113
T034	65	65	65	157	150	118	185	181	165	151	195	195	175	165	203	160	145	112	121	12	124	150	117	129	69	142	168	147	157	283	170	183	120
T011	56	56	56	190	166	110	248	244	239	189	221	223	241	236	213	218	208	187	192	19	118	120	110	118	70	179	221	198	188	170	339	251	166

T011	62	62	62	246	166	112	261	259	244	183	223	225	265	252	236	239	223	184	197	19	119	139	112	142	72	194	223	192	184	183	251	399	183
T034	55	55	55	172	132	88	173	169	165	154	171	169	176	171	156	158	150	123	146	13	95	98	88	95	58	140	156	142	113	120	166	183	267

Supplementary file 4: Novel variants of gyrA of 23 strains based on prediction on ARIBA.

ID	GYRA_A162S	GYRA_A709E	GYRA_D483E	GYRA_D596E	GYRA_D816E	GYRA_D856E	GYRAE409D	GYRA_E815D	GYRA_E817V	GYRA_E859V	GYRA_E862D	GYRA_E886D	GYRA_N842S	GYRA_N860T	GYRA_R875C	GYRA_S884L	GYRA_T818DEL	GYRA_T825_S826DEL	GYRA_T845A	GYRA_A223V
24977_8#292	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#293	NO	NO	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO	YES	NO	YES	YES	NO	NO
24977_8#294	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#310	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#313	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#314	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#323	NO	NO	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	YES	NO	NO	NO	YES	YES	YES	NO
24977_8#326	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#327	NO	NO	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO	YES	NO	YES	YES	NO	NO
24977_8#328	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO
24977_8#330	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#333	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	YES	NO	NO
24977_8#335	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#336	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	YES	NO	YES
24977_8#342	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#346	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#353	NO	NO	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	YES	NO	NO	NO	YES	YES	YES	NO
24977_8#356	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#368	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#371	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
24977_8#374	NO	NO	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	YES	NO	NO	NO	YES	YES	YES	NO
24977_8#380	NO	NO	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO	NO	NO	YES	NO	YES	YES	NO	YES
24977_8#383	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO

LANE	STUDY_ACCESSION	COUNTRY	YEAR	ST
15056_7#53	PRJEB2655	TANZANIA	2014	152
15056_7#58	PRJEB2655	TANZANIA	2014	152
ERR1143368	PRJEB11627	GERMANY	2011	152
ERR1143369	PRJEB11627	GERMANY	2010	152
ERR1143433	PRJEB11627	TANZANIA	2012	152
ERR1143437	PRJEB11627	TANZANIA	2012	152
ERR1143447	PRJEB11627	GABON	2011	152
ERR1143449	PRJEB11627	GABON	2011	152
ERR1143460	PRJEB11627	GABON	2011	152
ERR1143463	PRJEB11627	GABON	2012	152
ERR1143465	PRJEB11627	GABON	2012	152
ERR1143469	PRJEB11627	GABON	2012	152
ERR1143471	PRJEB11627	GABON	2012	152
ERR1143481	PRJEB11627	GABON	2013	152
ERR1143490	PRJEB11627	MOZAMBIQUE	2012	152
ERR1143492	PRJEB11627	MOZAMBIQUE	2012	152
ERR1195804	PRJEB8084	GERMANY	2014	152
ERR1213804	PRJEB12419	GAMBIA	2007	152
ERR1213807	PRJEB12419	GAMBIA	2008	152
ERR1213812	PRJEB12419	GAMBIA	2009	152
ERR1213817	PRJEB12419	GAMBIA	2009	152
7229_4#50	PRJEB2755	UK	1998	188
7915_6#22	PRJEB2944	UK	1998	188
9119_2#48	PRJEB2655	UNITED KINGDOM	2012	188
11641_2#13	PRJEB1915	USA	2010	188
11641_2#46	PRJEB1915	USA	2010	188
11641_2#68	PRJEB1915	USA	2009	188
12641_2#35	PRJEB1915	USA	2009	188
12641_2#75	PRJEB1915	USA	2010	188
12673_4#19	PRJEB1915	USA	2010	188
12673_5#80	PRJEB1915	USA	2004	188
12971_2#23	PRJEB1915	USA	2010	188
15056_7#55	PRJEB2655	TANZANIA	2014	188
17175_1#69	PRJEB9575	THAILAND	2015	188
17138_2#15	PRJEB9575	THAILAND	2015	188
17138_2#66	PRJEB9575	THAILAND	2015	188
17175_3#11	PRJEB9575	THAILAND	2015	188
17175_1#77	PRJEB9575	THAILAND	2015	188

4395_2#4	PRJEB2478	PORTUGAL	2006	188
ERR1143448	PRJEB11627	GABON	2011	6
ERR1143399	PRJEB11627	TANZANIA	2011	6
ERR1143403	PRJEB11627	TANZANIA	2011	6
ERR1143420	PRJEB11627	TANZANIA	2011	6
17175_2#4	PRJEB9575	THAILAND	2015	6
17175_2#16	PRJEB9575	THAILAND	2015	6
17175_2#17	PRJEB9575	THAILAND	2015	6
17138_2#39	PRJEB9575	THAILAND	2015	6
17138_2#45	PRJEB9575	THAILAND	2015	6
17138_2#48	PRJEB9575	THAILAND	2015	6
17138_2#59	PRJEB9575	THAILAND	2015	6
17175_1#84	PRJEB9575	THAILAND	2015	6
17175_1#85	PRJEB9575	THAILAND	2015	6
17175_1#47	PRJEB9575	THAILAND	2015	6
4351_2#8	PRJEB2096	AUSTRALIA: PERTH	1996	22
4351_8#1	PRJEB2096	AUSTRALIA: PERTH	2006	22
4351_8#10	PRJEB2096	AUSTRALIA: MELBOURNE	2006	22
4351_8#9	PRJEB2096	AUSTRALIA: PERTH	2006	22
4351_8#8	PRJEB2096	AUSTRALIA: MELBOURNE	2006	22
4395_5#1	PRJEB2096	AUSTRALIA: SYDNEY	2006	22
4351_8#11	PRJEB2096	AUSTRALIA: PERTH	2006	22
4351_8#2	PRJEB2096	AUSTRALIA: BRISBANE	2006	22
4351_8#3	PRJEB2096	AUSTRALIA: BRISBANE	2006	22
4351_8#5	PRJEB2096	AUSTRALIA: SYDNEY	2006	22
4351_8#7	PRJEB2096	AUSTRALIA: SYDNEY	2006	22
6437_8#7	PRJEB2510	AUSTRALIA	1999	22
7229_3#8	PRJEB2755	UK	1998	25
7748_6#30	PRJEB2756	UNITED KINGDOM: ENGLAND	2008	25
7748_6#4	PRJEB2756	UNITED KINGDOM: ENGLAND	2002	25
7748_6#27	PRJEB2756	UNITED KINGDOM: ENGLAND	2008	25
7748_6#24	PRJEB2756	UNITED KINGDOM: SCOTLAND	2007	25

7748_6#18	PRJEB2756	UNITED KINGDOM: ENGLAND	2006	25
11641_2#51	PRJEB1915	USA	2010	25
12625_6#13	PRJEB1915	USA	2010	25
12641_3#44	PRJEB1915	USA	2010	25
12673_1#66	PRJEB1915	USA	2009	25
12673_2#27	PRJEB1915	USA	2009	25
12673_2#28	PRJEB1915	USA	2009	25
12673_2#60	PRJEB1915	USA	2009	25
17175_2#54	PRJEB9575	THAILAND	2015	25
17175_2#60	PRJEB9575	THAILAND	2015	25
17175_3#47	PRJEB9575	THAILAND	2015	25
ERR1212598	PRJEB12240	UNITED KINGDOM	2008	25
ERR1213810	PRJEB12419	GAMBIA	2008	25
ERR1213816	PRJEB12419	GAMBIA	2009	25
7229_6#65	PRJEB2755	UK	1998	789
ERR1195979	PRJEB8084	GERMANY	2015	789
ERR1050518	PRJEB11281	DENMARK	2014	398
ERR1699818	PRJEB12818	USA	2011	398
ERR1699820	PRJEB12818	USA	2011	398
ERR1753508	PRJEB18560	DENMARK	2013	398
ERR1040953	PRJEB11177	UNITED KINGDOM	2011	398
ERR1699816	PRJEB12818	USA	2007	398
ERR1699817	PRJEB12818	USA	2011	398
6133_1#11	PRJEB2478	FRANCE	2006	398
ERR1050515	PRJEB11281	DENMARK	2013	398
ERR1050520	PRJEB11281	DENMARK	2013	398
ERR1429006	PRJEB14187	FINLAND	2015	398
ERR1682039	PRJEB12552	NEW ZEALAND	2015	398
ERR493469	PRJEB6236	DENMARK	2008	398
ERR593597	PRJEB7089	GERMANY	2014	398
ERR593602	PRJEB7089	GERMANY	2014	398
ERR593609	PRJEB7089	GERMANY	2014	398
8113_4#24	PRJEB2755	UK	1998	398
ERR1050516	PRJEB11281	DENMARK	2013	398
ERR1050517	PRJEB11281	DENMARK	2013	398
ERR1050519	PRJEB11281	DENMARK	2014	398
ERR1429004	PRJEB14187	FINLAND	2015	398
ERR1429005	PRJEB14187	FINLAND	2015	398

ERR1682013	PRJEB12552	NEW ZEALAND	2013	398
ERR1682033	PRJEB12552	NEW ZEALAND	2015	398
ERR1682037	PRJEB12552	NEW ZEALAND	2015	398
ERR1682038	PRJEB12552	NEW ZEALAND	2015	398
ERR493447	PRJEB6236	DENMARK	2007	398
ERR493466	PRJEB6236	DENMARK	2007	398
ERR593112	PRJEB7089	GERMANY	2013	398
ERR593115	PRJEB7089	GERMANY	2013	398
17262_2#28	PJEB9644	GHANA	2012	152
17262_2#29	PJEB9644	GHANA	2012	152
17262_2#40	PJEB9644	GHANA	2012	152
SRR3047795	PJNA306753	USA	2015	188
SRR3047865	PJNA306753	USA	2015	188
SRR3047885	PJNA306753	USA	2015	188
SRR3047953	PJNA306753	USA	2015	188
SRR3048015	PJNA306753	USA	2015	188
SRR3048017	PJNA306753	USA	2015	188

Supplementary file 6: Public genomes showing lane numbers in the Sanger cluster together with accession study project numbers