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.